



UNIVERSIDAD DE CHILE
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
DEPARTAMENTO DE INGENIERÍA ELÉCTRICA

**TRANSFORMACIÓN DIGITAL A NIVEL CORPORATIVO MEDIANTE LA
REALIZACIÓN DE PROYECTOS BASADOS EN ANALYTICS Y MACHINE
LEARNING**

MEMORIA PARA OPTAR AL TÍTULO DE INGENIERO CIVIL ELÉCTRICO

PABLO ANDRÉS TRONCOSO PONCE

PROFESOR GUÍA:
NICOLÁS TORO SALAZAR

MIEMBROS DE LA COMISIÓN:
IGNACIO MENARES JIMÉNEZ
ANDRÉS CABA RUTTE

Este trabajo ha sido parcialmente financiado por:
Falabella S.A.

SANTIAGO DE CHILE
2022

RESUMEN DE LA MEMORIA PARA OPTAR AL
TÍTULO DE INGENIERO CIVIL ELÉCTRICO
POR: PABLO ANDRÉS TRONCOSO PONCE
FECHA: 2022
PROF. GUÍA: NICOLÁS TORO SALAZAR

TRANSFORMACIÓN DIGITAL A NIVEL CORPORATIVO MEDIANTE LA REALIZACIÓN DE PROYECTOS BASADOS EN ANALYTICS Y MACHINE LEARNING

El presente trabajo se refleja como una iniciativa pionera en la gerencia de finanzas de Falabella S.A., donde mediante el diseño e implementación de 3 proyectos de transformación digital basados en analítica y machine learning, se busca generar casos de éxito de uso de tecnologías digitales para la mejora de procesos y el aporte en la transformación digital de la gerencia.

Primero se realiza una revisión bibliográfica para la comprensión de la magnitud de la transformación digital, las tecnologías digitales mayormente involucradas en las empresas como la automatización de procesos, servicios de cloud computing, elementos desde la analítica avanzada y aplicaciones propias de la inteligencia artificial como el machine learning y los algoritmos predictivos de Prophet Y NeuralProphet.

Se procede con una metodología que permite el desarrollo de los 3 proyectos de transformación digital, abarcando desde la identificación de necesidades a todas las etapas del trabajo con datos con principios desde la ciencia de datos. De esta manera se obtuvieron los resultados para los 3 proyectos, los cuales generaron predicciones de series de tiempo; ya sea venta neta en Tottus Chile, venta en Sodimac y deuda acumulada en Falabella Inmobiliaria con bajas tasas de error: menores al 15% para las predicciones diarias y menores al 10% en las predicciones mensuales, exceptuando casos puntuales donde el desempeño supera estos valores.

A su vez, los modelos desarrollados lograron impactos y beneficios relevantes en las unidades de negocio, incorporando en la práctica usos de tecnologías desde la programación de tareas y automatización de estas, uso de servicios de la nube como bigquery de google cloud platform o algoritmos de machine learning que permitieron mejorar procesos financieros involucrados y expandir los desarrollos a más procesos, ya sea dentro de las mismas unidades de negocio o a otras dentro de Falabella S.A., sirviendo como casos de éxito en la transformación digital de la gerencia de finanzas corporativa.

*A mi vieja querida,
ejemplo de esfuerzo y sacrificio
que demuestra que todo se puede.*

Agradecimientos

En primer lugar, agradecer a mi familia por su apoyo, su cariño incondicional día a día y sus sacrificios para que no faltase nada en la casa. A mi madre por ser un ejemplo de esfuerzo y cariño, por ser esa persona afable en quien siempre podía reposar después de una día agotador. A mi viejo, por mañosearme con comida rica para las noches de estudio, por esas enseñanzas de vida y por darme ese espíritu crítico para la vida. A mi hermanito el molesto, por ser ese apoyo desde las sombras, por su preocupación a su manera, por su ánimo y optimismo para terminar la carrera.

Agradecer a mi Cami, por el amor y gentileza que me has dado, por mañosearme cuando me veía cansado, por motivarme en los días de trabajo, por las risas y por cambiar de la noche a la mañana este 2021. Solo agradecer las casualidades y que venga el futuro.

Muchas gracias a todas las amistades durante mi paso en la universidad, a los cabros del colegio que nos mantuvimos en plan común, a los “K” por los carretes en las bancas de 850 y mención especial al Nico y la Clau, por ser las ratas esos primeros años de eléctrica, por las tarde/noches haciendo informes y por la compañía de esos años. Los quiero caleta.

Agradecer a los chicos y chicas de Sumariados por las risas, las tardes de estudio en la bibliodíe, los llantos por SEP y los carretes de halloween de la Malu. A los cabros de eléctrica por las risas, los trolleos, las noches de lolazos o 7days y los carretes en el depa embrujado del Diego. A las bbs de eléctrica: Cata y Francys, por ser ese apoyo hard en la escuelita online y por las tardes de conversas en discord. Al Nico, Claudito, Bryan, Tomi y Pep, por ser ese núcleo el último tiempo que espero nos siga reuniendo de aquí a cuando seamos viejos recordando historias del pasado.

Agradecer también a la gente Electrotutores, un espacio que me permitió crecer y dar mi granito de arena en el departamento. A los chiqs con que trabajé en redes beauchef, amor y odio a un proyecto que nos encantaba pero nos sacaba canas verdes.

Por último y no menor, agradecer a los miembros de la comisión; a Nicolás por la oportunidad y confianza de trabajar juntos y permitir desarrollarme en un área que me entusiasma. A Ignacio por todo el apoyo entregado en estos casi 7 meses de trabajo, desde las dudas día a día a la gentileza en las conversaciones de Teams. Y Andrés, profesor durante la carrera, jefe en Electrotutores y una persona a la cual siempre se podía acudir en búsqueda de consejos, simplemente agradecer tu disposición, los consejos y los apañes. A los tres gracias sinceras por su ayuda, por las risas, por el trabajo y por ser las buenas personas que son.

Tabla de Contenido

1. Introducción	1
1.1. Motivación	1
1.2. Objetivos	2
1.2.1. Objetivo General	2
1.2.2. Objetivos Específicos	2
1.3. Alcances	2
1.4. Estructura del documento	3
2. Marco teórico	4
2.1. Transformación digital	4
2.1.1. Ejes estratégicos de la transformación digital	5
2.1.2. Transformación digital en finanzas	6
2.1.3. Casos de éxito de transformación digital	7
2.1.3.1. Maerks	7
2.1.3.2. Empresa multinacional	7
2.1.3.3. Banco multinacional	8
2.2. Tecnologías y disciplinas relacionadas a la transformación digital	8
2.2.1. Automatización robótica de procesos	9
2.2.2. Tecnologías de la nube	9
2.2.3. Ciencia de datos	9
2.2.4. Analítica	11
2.3. Metodologías para proyectos de ciencia de datos	11
2.4. Inteligencia artificial	13
2.5. Aprendizaje de máquinas	14
2.6. Algoritmos de Machine Learning	16
2.6.1. Técnicas de regresión	17
2.6.1.1. Regresión lineal	17
2.6.2. Pronóstico de series de tiempo	18
2.6.2.1. Prophet	19
2.6.2.2. NeuralProphet	20
2.7. Validación y métricas de error	22
3. Metodología	24
3.1. Propuesta metodológica	24
3.2. Implementación	27
3.2.1. Identificación de necesidades y oportunidades de proyectos en 3 unidades de negocios	27

3.2.2.	Desarrollo de proyectos	28
3.2.2.1.	Proyecto: Predicción de ventas de Tottus	28
3.2.2.2.	Proyecto: Predicción de ventas diarias y mensuales en Sodimac Latam	34
3.2.2.3.	Proyecto: Predicción de deuda acumulada mensual para Falabella Inmobiliaria	40
4.	Resultados y Análisis	46
4.1.	Predicción de ventas para Tottus Chile	46
4.1.1.	Impactos y beneficios del proyecto	48
4.2.	Predicción de ventas para Sodimac Latam	52
4.2.1.	Pronóstico de venta diaria	52
4.2.2.	Pronóstico de venta mensual	62
4.2.3.	Impactos y beneficios del proyecto	71
4.3.	Predicción de deuda acumulada para Falabella Inmobiliaria	73
4.3.1.	Pronóstico de deuda acumulada del negocio	73
4.3.2.	Pronóstico de deuda acumulada por sociedades	75
4.3.3.	Impactos y beneficios del proyecto	79
5.	Conclusiones	81
5.1.	Trabajo Futuro	82
	Bibliografía	84

Índice de Tablas

3.1.	VARIABLES ESTADÍSTICAS DE LAS VENTAS NETAS	29
3.2.	Ejemplo de estructura de data vías a modelo Prophet	32
3.3.	Revisión estadística de ventas en países de Sodimac	35
3.4.	Ejemplo de estructura de data para modelos en Sodimac	38
3.5.	Parámetros para modelos de predicción de ventas en base a Prophet	40
3.6.	Parámetros para modelos de predicción de ventas en base a NeuralProphet	40
3.7.	Ejemplo de estructura de data para modelos en Falabella Inmobiliaria	43
3.8.	Parámetros para modelos de predicción de deuda acumulada en base a Prophet	45
3.9.	Parámetros para modelos de predicción de deuda acumulada en base a Neural-Prophet	45
4.1.	Métricas de error en predicción mensual de ventas Tottus Chile	47
4.2.	Comparación métricas de error predicción de ventas mensuales de acuerdo a proceso actual	48
4.3.	Métricas de error en predicción diaria de ventas Tottus Chile	50
4.4.	Comparación métricas de error predicción de ventas mensuales para local de Los Dominicos	51
4.5.	Métricas predicción ventas diarias Chile	53
4.6.	Métricas predicción ventas diarias Colombia	55
4.7.	Métricas predicción de ventas diarias - Argentina	56
4.8.	Métricas predicción ventas diarias Brasil	57
4.9.	Métricas predicción ventas diarias Uruguay	58
4.10.	Métricas predicción ventas diarias México	59
4.11.	Métricas predicción ventas diarias Perú	60
4.12.	Métricas predicción ventas mensuales Chile	63
4.13.	Métricas predicción ventas mensuales Colombia	64
4.14.	Métricas predicción ventas mensuales Argentina	65
4.15.	Métricas predicción ventas mensuales Brasil	66
4.16.	Métricas predicción ventas mensuales Uruguay	67
4.17.	Métricas predicción ventas mensuales México	68
4.18.	Métricas predicción ventas mensuales Perú	69
4.19.	Métricas predicción deuda acumulada total negocio F. Inmobiliaria	74
4.20.	Métricas predicción deuda acumulada Sociedad C069	76
4.21.	Métricas predicción deuda acumulada Sociedad C106	78

Índice de Figuras

2.1.	Etapas generales del proceso de transformación digital [1].	5
2.2.	Principales enfoque de Transformación digital en finanzas [3].	6
2.3.	Áreas de la ciencia de datos [10].	10
2.4.	Marco de referencia de inteligencia artificial, machine learning y ciencia de datos [10].	10
2.5.	Etapas de metodología KDD. [14].	12
2.6.	Etapas de metodología SEMMA [15].	12
2.7.	Etapas de metodología CRISP-DM [16].	13
2.8.	Esquema representativo de aprendizaje supervisado [19].	14
2.9.	Esquema representativo de aprendizaje no supervisado [19].	15
2.10.	Esquema representativo de aprendizaje reforzado [19],	15
2.11.	Esquema representativo de redes neuronales, base del deep learning [21].	16
2.12.	Ejemplos de regresión lineal [23].	17
2.13.	Marco referencia modelos de series de tiempo [10].	18
2.14.	Representación modelo de timeseries forecasting[10].	18
2.15.	Ejemplo de funcionamiento de AR-NET [26].	21
3.1.	Diagrama de metodología propuesta a seguir.	24
3.2.	Ventas diarias Tottus Chile	30
3.3.	Ventas mensuales Tottus Chile	30
3.4.	Costos mensuales Tottus Chile	31
3.5.	Unidades vendidas mensuales Tottus Chile	31
3.6.	Ventas diarias Sodimac Latinoamérica	35
3.7.	Ventas diarias Sodimac AR-BR-PE-MX-UY	36
3.8.	Ventas diarias Sodimac BR-PE-MX-UY	36
3.9.	Ventas diarias por país - Sodimac.	37
3.10.	Deuda acumulada real para todo el negocio de F. Inmobiliaria	42
3.11.	Deuda acumulada real para las sociedades de F. Inmobiliaria	43
4.1.	Predicción ventas mensuales para Tottus Chile	47
4.2.	Predicción ventas mensuales para Tottus Chile 2021	47
4.3.	Predicción ventas diarias Tottus Chile	49
4.4.	Predicción ventas diarias Tottus Chile 2021	50
4.5.	Predicción ventas mensuales para local de Los Dominicos Tottus 2021	51
4.6.	Predicción ventas diarias - Sodimac Chile 2021	53
4.7.	Predicción ventas diarias - Sodimac Colombia	54
4.8.	Predicción ventas diarias - Sodimac Argentina	55
4.9.	Predicción ventas diarias - Sodimac Brasil 2021	56
4.10.	Predicción ventas diarias - Sodimac Uruguay	57
4.11.	Predicción ventas diarias - Sodimac México	58

4.12.	Predicción ventas diarias - Sodimac Perú	59
4.13.	Predicción ventas diarias Sodimac Latam 1	61
4.14.	Predicción ventas diarias Sodimac Latam 2	61
4.15.	Predicción ventas diarias Sodimac Latam 3	62
4.16.	Predicción ventas mensuales - Sodimac Chile	62
4.17.	Predicción ventas mensuales - Sodimac Colombia	64
4.18.	Predicción ventas mensuales - Sodimac Argentina	65
4.19.	Predicción ventas mensuales - Sodimac Brasil	66
4.20.	Predicción ventas mensuales - Sodimac Uruguay	67
4.21.	Predicción ventas mensuales - Sodimac México	68
4.22.	Predicción ventas mensuales - Sodimac Perú	69
4.23.	Predicción ventas mensuales Sodimac Latam 1	70
4.24.	Predicción ventas mensuales Sodimac Latam 2	71
4.25.	Predicción ventas mensuales Sodimac Latam 3	71
4.26.	Predicción deuda acumulada mensual para todo el negocio de Falabella Inmobiliaria	74
4.27.	Resultados de predicción de deuda acumulada mensual para todo el negocio de Falabella Inmobiliaria	75
4.28.	Predicción deuda acumulada mensual para sociedad C069 de Falabella Inmobiliaria	76
4.29.	Resultados de predicción deuda acumulada mensual para sociedad C069 de Falabella Inmobiliaria	77
4.30.	Predicción deuda acumulada mensual para sociedad C106 de Falabella Inmobiliaria	78
4.31.	Resultados de predicción deuda acumulada mensual para sociedad C106 de Falabella Inmobiliaria	79

Capítulo 1

Introducción

1.1. Motivación

El mundo empresarial está constantemente frente a adaptaciones en su entorno comercial, donde el rol ejercido por las tecnologías digitales ha acelerado los cambios de los últimos años en la considerada cuarta etapa de la revolución industrial.

La verticalidad general de mercados se ha visto quebrantada frente a la entrada de nuevos participantes con la insignia del uso de tecnologías digitales tanto en sus procesos internos como en su oferta a terceros, tal cual lo hace Amazon, Alibaba. Estas empresas trascienden de su rol en el mercado minorista para atacar otras industrias bajo el principio de “aprenderemos más rápido de la complejidad de otras industrias, que lo que aprenderán las empresas de esas industrias sobre tecnologías digitales” [1].

A fin de cuentas, la modernización relacionada a las tecnologías ha irrumpido los mercados y ha presionado a las empresas tradicionales a adaptarse a esta era digital en sus distintas áreas, buscando evitar la obsolescencia y generar nuevos enfoques que permita captar sus dividendos. Y estos beneficios son claros, donde el Foro Económico Mundial estima que la transformación digital impulse 100 billones de dolares a la economía mundial para 2025 [2].

Falabella de la misma manera se enfrenta a la problemática de verse desplazado por competidores locales como Mercado Libre, quienes desde la tecnología mejoran sus procesos y prestación de servicios hacia la toma de un rol protagónico en el mercado minorista, y continuamente incursionando en otros negocios. Falabella al ser una empresa tradicional posee una gran cantidad de procesos anticuados y tediosos que requieren adaptarse a la era digital, en especial en el área de finanzas caracterizada por su rol de velar por la salud financiera de la compañía.

En concreto, Falabella S.A y su gerencia de Finanzas presenta desventajas competitivas, requiriendo un proceso de modernización de la mano de adoptar tecnologías digitales en sus procesos, adoptar un enfoque “data drive” para las decisiones y todo elemento que agregue valor a sus procesos en una visión de transformación digital de la gerencia que permita ser el camino para que Falabella pueda mantenerse competitiva en el mundo empresarial.

1.2. Objetivos

1.2.1. Objetivo General

El objetivo de la presente memoria consiste en diseñar e implementar 3 proyectos de transformación digital basados en analítica y machine learning, que generen mejoras en procesos financieros de las gerencias de finanzas de Falabella S.A. y sus unidades de negocios, aportando de esta manera en la transformación digital de la compañía.

1.2.2. Objetivos Específicos

- Identificar y caracterizar necesidades en 3 unidades de negocios, vinculando estas hacia soluciones mediante técnicas de Machine Learning.
- Desarrollar activos analíticos en cada proyecto a partir del uso provechoso de los datos disponibles en las unidades de negocio.
- Seleccionar e implementar algoritmos de machine learning flexibles e interpretables en su desarrollo y resultados generados para cada uno de los proyectos.
- Definir indicadores que permitan medir beneficios e impactos en cada uno de los proyectos y sus procesos financieros involucrados.

1.3. Alcances

- Los proyectos realizados no consideran implementación comercial ni medición cuantitativa del dinero, pues la evaluación práctica de esto consideraría historia financiera post cierre del 2021 o 2022.
- Los algoritmos de machine learning trabajados se enfocarán principalmente en predicciones sobre series de tiempo, esto al ser el principal tipo de dato involucrado en procesos financieros, no abarcando otro tipo de algoritmos de machine learning.
- Los modelos de machine learning a implementar se enfocan en modelos interpretables que se puedan explicar sus resultados con su metodología interna y no sean cajas negras incomprensibles.
- Falabella S.A. y sus unidades de negocio pondrán a disposición toda la información a utilizar en la realización de la memoria, dando el carácter de receptores y consumidores de datos.

1.4. Estructura del documento

Esta memoria está compuesta por cinco capítulos, comenzando por la introducción a la temática desde la motivación que impulsó el trabajo realizado, la presentación de objetivos generales y específicos para finalizar con los alcances prácticos del trabajo realizado.

En el segundo capítulo se presenta el marco teórico de los tópicos relacionados a la memoria desarrollada y que permiten su comprensión a cabalidad, haciendo una pasada por elementos de transformación digital y sus etapas, tecnologías digitales involucradas y conceptos relacionados a la inteligencia artificial, ciencia de datos, machine learning y los principales algoritmos involucrados en el trabajo desarrollado.

En el capítulo 3 se presenta la propuesta metodológica del trabajo realizado, describiendo sus etapas e implementación de ellas para cada uno de los proyectos realizado, permitiendo ahondar sobre los datos utilizados, preprocesamiento de estos y algoritmos de machine learning implementados.

En el capítulo 4 se exponen los resultados obtenidos para cada proyecto, analizando estos de acuerdo a la información financieras de las unidades de negocio, los distintos algoritmos implementados, y en caso que corresponda, comparando con los procesos actuales de las unidades de negocio. Además, se incluyen mediciones de impactos y beneficios de los proyectos, tanto en la precisión de los algoritmos como en aspectos de mejora de proceso, adopción de tecnologías y cualquier otro elemento en la línea de la transformación digital.

Para finalizar, en el capítulo 5 se señalan las principales conclusiones del trabajo, los impactos de los 3 proyectos desarrollados y sus resultados obtenidos, evaluando el cumplimiento de objetivos tanto general como específicos. Junto a esto se deslizan recomendaciones de trabajo futuro enfocado en los algoritmos predictivos implementados y en aspectos que logren cimentar la transformación digital dentro de la organización.

Capítulo 2

Marco teórico

2.1. Transformación digital

La transformación digital se entiende como un proceso amplio de integración de tecnologías digitales en el ADN de los procesos de negocios de una empresa u organización para satisfacer los emergentes requerimientos del mercado y sus clientes en la era digital [1].

Su necesidad nace frente a la relevancia que ha tomado lo digital en los últimos años, destacándose en tres líneas principalmente:

- La vasta disponibilidad de datos, el manejo masivo de datos y las tecnologías digitales emergentes como la inteligencia artificial, el cloud computing, el internet de las cosas o la robótica.
- La revolución en la competencia donde empresas jóvenes y digitales como Amazon, Alibaba o Facebook han adquirido un rol protagónico en el mercado.
- El cambio de preferencias de los clientes cada vez más informados, exigentes, activos y preferentes al mundo digital donde los dispositivos móviles, la masificación de las redes de telecomunicaciones y las redes sociales han jugado un rol importante.

Esta transformación es propiamente multidisciplinaria dado que considera cambios de estrategia, cultura organizacional y utilización de tecnologías digitales con el objetivo de innovar y generar valor para la empresa en la considerada cuarta revolución industrial.

De manera general, la transformación digital se consideran 3 fases principalmente: la digitación, la digitalización y finalmente la transformación digital pura como se observa en la figura 2.1 [1].

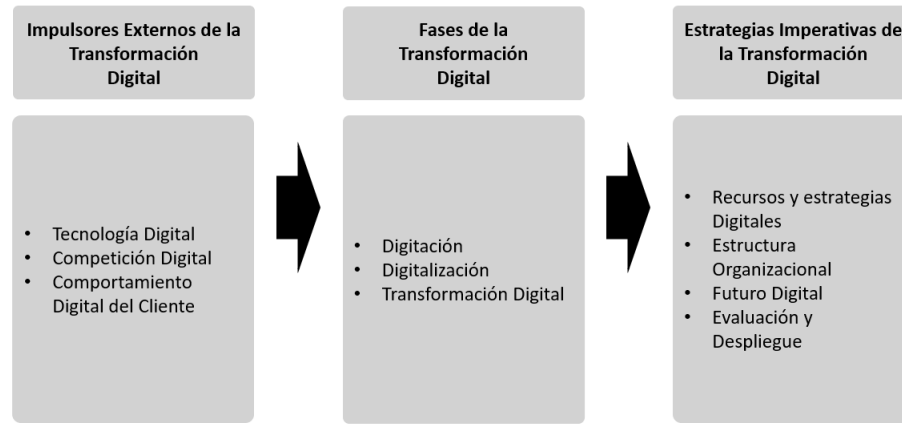


Figura 2.1: Etapas generales del proceso de transformación digital [1].

- **Digitación:** Del inglés “digitization”, esta fase corresponde al paso de información de un formato analógico a uno digital con el poder de utilizar esta mediante la informática y la computación. En la práctica hace referencia a pasar del papel al computador, la modificación de la forma de utilizar dicha información y la automatización de procesos.
- **Digitalización:** Describe a la etapa enfocada en utilizar tecnologías digitales para modificar y hacer más eficiente los procesos, las formas de trabajo y la creación de valor de la empresa u organización. Esto en la práctica se aplicaba en la utilización de tecnologías de la información para la comunicación con clientes, de softwares para la optimización de procesos como contabilidad o logística, entre otros.
- **Transformación digital:** Finalmente procede la etapa de la transformación la cual describe todo cambio en la empresa u organización que mediante la utilización de tecnologías digitales, conduce a la generación de nuevos procesos y modelos de negocios. Este elemento se traduce en un cambio de enfoque a la cultura de decisiones basadas en datos, mejor conocida de la expresión del inglés “data driven”.

2.1.1. Ejes estratégicos de la transformación digital

En el desarrollo de la transformación digital existen elementos clave donde este proceso debe impactar sobre el status-quo, aprovechar los recursos involucrados y permitir la evolución de la organización para ser competitiva en un mercado cada vez más digital [1].

A continuación, se mencionan 4 ejes claves sobre los cuales se basan gran parte de las hojas de ruta y marcos de referencias que utilizan consultoras internacionales para guiar este cambio digital [1].

- **Recursos y estrategias digitales:** La cual hace referencia al uso de los recursos digitales de la organización con foco en la redefinición de procesos y tareas de manera de aprovechar los datos disponibles y el trabajo con estos para crear valor mediante diversos análisis.

- **Estructura organizacional:** Corresponde a una de las aristas más relevantes en las transformaciones digitales asociada al proceso de adaptación y modificación de status-quo de las organizaciones, impactando en los equipos y recursos humanos hacia un enfoque con mayor flexibilidad, agilidad y autonomía que permita dinamizar la aplicación de cambios tecnológicos y explotar las oportunidades de negocio desde las tecnologías digitales.
- **Futuro digital:** Consiste en el núcleo de transformación asociada a como se rediseñan, reconsideran y refocalizan los servicios, productos y focos de negocio de la organización hacia un mundo digital, tal cual lo hace Google o Amazon innovando en industrias no digitales y constantemente probando nuevos productos y servicios.
- **Evaluaciones e implementaciones:** Finalmente, el último eje estratégico se orienta hacia la consideración de nuevos enfoques de medir resultados e impactos asociados a la incorporación de tecnologías digitales, haciendo hincapié en evaluar cuantitativamente elementos de rentabilidad y eficiencia junto a cualitativamente rediseño de procesos o elementos estratégicos de la organización.

2.1.2. Transformación digital en finanzas

La realización y camino de la transformación digital posee la capacidad de adaptarse a la envergadura y estructura organizacional de las empresas, en especial de grandes empresas con gran cantidad de gerencia y subdivisiones. De esta manera se permite enfocar las transformaciones en áreas específicas y no verse entorpecido en el global.

En específico sobre las áreas de finanzas, las cuales requieren atención por su rol con la información comercial, transaccional y a fin de cuentas de la salud financiera de la organización, se permite enfocar la transformación digital en modernizar sus procesos incorporando tendencias tecnológicas, uso de analítica y modelos desde el machine learning que permita repensar el quehacer financiero [3].

Concretamente, el consenso diagnosticado por consultoras como Mckinsey and Company muestra que los CFOs y agentes involucrados en las áreas financieras creen que la automatización de procesos financieros, la visualización en tiempo real, la analítica enfocada en decisiones de operaciones financieras y la analítica enfocada a descubrir oportunidades de negocio no explotadas son el camino a futuro en las finanzas de las organizaciones.

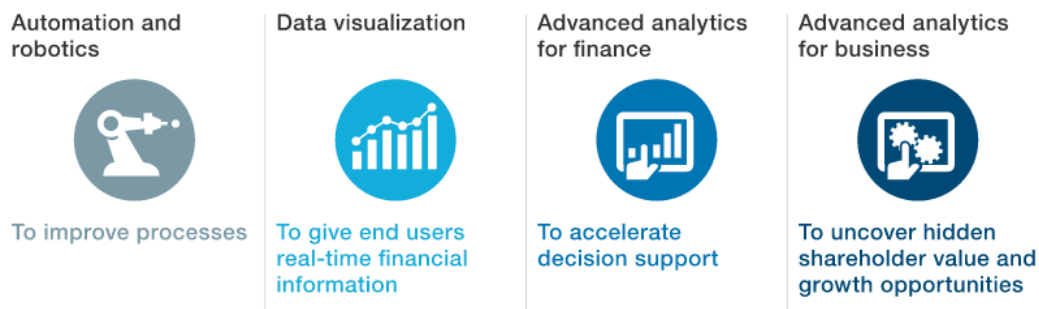


Figura 2.2: Principales enfoque de Transformación digital en finanzas [3].

2.1.3. Casos de éxito de transformación digital

Empresas que se han aventurado en los procesos de transformación digital cada vez son más frecuentes, considerando el impacto que tienen las tecnologías digitales en el comercio y la fuerte competencia de empresas nativas digitales como Amazon o Google.

A continuación, se presentan un resumen de casos de éxito de empresas tradicionales que se han sumergido en el proceso de transformación digital, en especial, con casos relacionados a áreas de finanzas.

2.1.3.1. Maerks

AP Moller - Maerks, es una empresa dedicada a la logística (rutas comerciales, terminales, remolcadores y manufactura de contenedores) con sede en Copenhague, Dinamarca con 76.000 empleados en 130 países e ingresos el 2019 de 41mm USD [4].

Esta empresa de un rubro no digital inició su proceso de transformación digital el 2015 buscando llevar los datos a la toma de decisiones mediante la experimentación a baja escala y de la combinación de datos financieros con no financieros.

Específicamente en el equipo de finanzas, se generaron distintas estrategias para la creación de valor a partir de elementos de la transformación digital:

- Rediseño de perspectivas del trabajo del área para la generación de valor, incorporando elementos digitales y tangibles.
- Re-enfoque de talento del equipo, incorporando y capacitando capital humano.
- Desarrollo de herramientas para la planificación operativa a partir de herramientas de datos, automatización de procesos y algoritmos predictivos de machine learning.
- Generación de proyectos a baja escala, con metodología ágil de manera de aprender rápido, evaluar impactos, generar retroalimentación y escalar casos que permitieran crear valor.
- Generación de modelos para la evaluación de sensibilidad de presupuestos de adquisiciones con reducción de 1.4 billones de dólares.

2.1.3.2. Empresa multinacional

Un conglomerado multinacional, cuyo nombre de protege en el reporte de PWC, poseía operaciones en múltiples industrias y regiones con un crecimiento los últimos años a través de adquisiciones agresivas, donde estas adquisiciones comenzaron a generar problemas de planificación de recursos, falta de estandarización e integración de acuerdo con sus procesos separados [5].

El alto costo diagnosticado de estos problemas motivó a la empresa a realizar una transformación financiera que permitiera abordar sus problemas estructurales a partir de aprovechar las tecnologías digitales. Esta transformación se realizó mediante el apoyo de una consultora internacional con actividades claves como:

- Automatización robótica de tareas (RPA) con foco en tareas repetitivas y transversales.
- Creación de un lago de datos estratégico que permitiera unificar las fuentes de datos de sus distintos negocios y adquisiciones.
- Estandarización de procesos entre negocio y adquisiciones, y de sistemas de planificación de recursos.
- Cambios de enfoques estratégicos que involucrara a su capital humano, su estructura organizacional y su cultura de trabajo.

Esta transformación masiva iniciada hace más de cinco años generó impactos evaluables relevantes como reducciones de costos de procesos financieros, consolidación de la digitalización en sus procesos, incremento de precisión de modelos de ingreso y riesgo financieros mediante herramientas de analítica, entre otros.

2.1.3.3. Banco multinacional

Uno de los bancos más grandes del mundo dentro de su ambiciosa transformación digital estaba actualizando su modelo operativo y su panorama tecnológico. Uno de los principales obstáculos era la dependencia del juicio humano y la falta de herramientas analíticas que permitiesen ayudar en la generación de proyecciones financieras [6].

Mediante el apoyo de una consultora, el área de finanzas logró rediseñar su proceso operativo en la generación de presupuestos de ingresos sin intereses basado en tarifas. Esto se realizó por medio de la utilización de modelos basados en inteligencia artificial, capacitación del área en herramientas digitales y el liderazgo desde el director de finanzas.

Este caso específico dentro de la transformación digital de un área de un banco importante logró resultados cuantificables como disminución de horas hombre dedicadas a la generación de presupuestos o las reducciones de los proyecciones de ingresos sin tarifas con impacto relevante en las finanzas del banco.

2.2. Tecnologías y disciplinas relacionadas a la transformación digital

Dentro de las distintas fases de la transformación digital y los procesos que involucra, el uso de tecnologías digitales juega un rol clave en el cambio de enfoque de tareas y procesos de las empresas u organizaciones.

A continuación, se relatan brevemente un grupo de tecnologías y disciplinas frecuentemente vinculadas con este cambio cultural que se destacan por su impacto y potencial a futuro.

2.2.1. Automatización robótica de procesos

La automatización robótica de procesos, o RPA de las siglas del inglés, consiste en el uso de procesos robóticos para imitar tareas altamente rutinarias y repetitivas realizadas por humanos con el fin de automatizarlas mediante uso de software y tecnologías digitales [7][8].

El impacto del RPA posee varias dimensiones, por un lado, se libera capital humano de la realización de tareas repetitivas lo que permite su reconversión a tareas que generen mayor valor con enfoques del área digital. De la misma manera, la automatización permite reducir los errores humanos en tareas repetitivas, mejorando la precisión y cumplimiento de procesos dentro de las organizaciones.

Ejemplos de usos del RPA se ven en la línea de la automatización de reportes, procesos de factura, elaboración de informes periódicos o el tratamiento base de datos como extracciones, actualizaciones o validaciones de información. Industrias con alta tasa de tareas repetitivas se ven altamente beneficiadas por el RPA como la industria financiera, de seguros o la venta minorista.

2.2.2. Tecnologías de la nube

Las tecnologías de nube de datos, o del inglés “cloud computing”, corresponde a todo el conjunto de infraestructura, tecnologías y servicios de recursos informáticos que permiten el acceso remoto, almacenamiento de información y procesamiento de datos por medio de internet, sirviendo de esta manera, como una alternativa a su realización local.

Los principales potenciales que adquiere la computación en la nube es la flexibilidad, agilidad y escalabilidad que permiten su servicio. Esto genera impactos en operaciones para la organización y permiten afianzar el cumplimiento criterios de seguridad sobre los datos que poseen las empresas sobre sus operaciones y clientes [9]. Los principales proveedores de cloud computing hoy en día en el mundo son Amazon, Google, Microsoft y Alibaba.

Ejemplos de uso práctico de las tecnologías de computación en la nube se ven en el almacenamiento y procesamiento de grandes volúmenes de datos, analítica en línea con bigdata, implementación de modelos en línea, generación de herramientas analíticas, entre otros.

2.2.3. Ciencia de datos

La ciencia de datos corresponde a la disciplina que combina las áreas de las matemáticas, estadística, ciencia y método científico, programación, analítica e inteligencia artificial para la extracción de conocimiento y valor a partir de datos [10], convirtiéndola en uno de los campos multidisciplinarios de mayor relevancia los últimos años.

Este campo abarca desde la extracción y preparación de datos, el uso de analítica para encontrar correlaciones significativas en grandes conjuntos de datos, la extensión de resultados para su visualización y la traducción de estos a conclusiones relevantes. De esta manera se busca descubrir nuevas preguntas y transformar estas a estrategias procesables por la organización.

Por su naturaleza, la ciencia de datos resulta una disciplina relevante para los procesos de transformación digital, donde los científicos de datos poseen un perfil atractivo para las organizaciones.

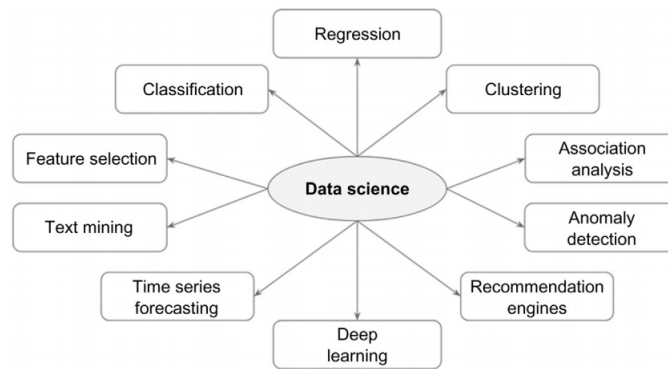


Figura 2.3: Áreas de la ciencia de datos [10].

Debido a su carácter multidisciplinario, el concepto de ciencia de datos suele confundirse con otros conceptos y tecnologías como la inteligencia artificial, el aprendizaje de máquinas o el deep learning, sin embargo, el carácter de “ciencia” es su principal diferenciador al ser un campo que mediante elementos de distintas áreas utiliza el método científico para la resolución de problemas.

Una clara diferenciación de estos conceptos se logra apreciar en la figura 2.4.

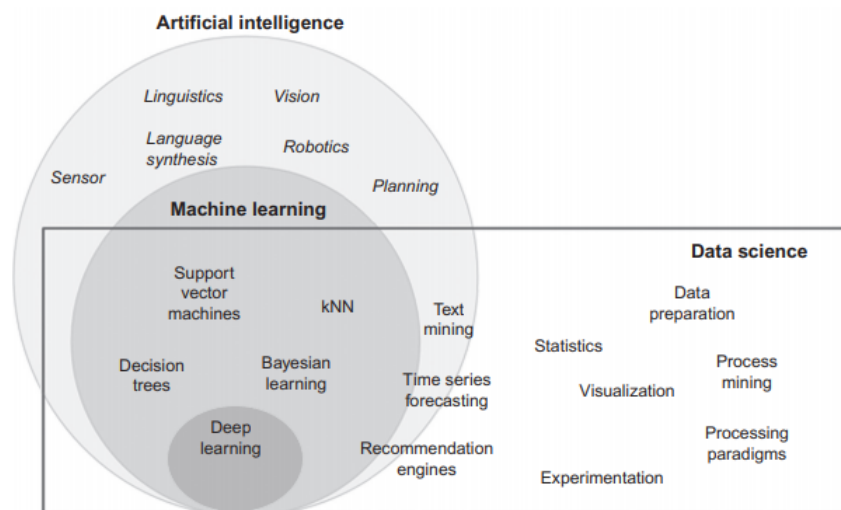


Figura 2.4: Marco de referencia de inteligencia artificial, machine learning y ciencia de datos [10].

2.2.4. Analítica

La analítica, o analytics del término en inglés, consiste en un subcampo de la ciencia de datos que busca descubrir e identificar patrones destacables en los datos y generar conocimientos para la toma de decisiones.

La analítica se basa en la unión de áreas de la matemática aplicada, la estadística, modelos predictivos y técnicas del machine learning aplicados sobre datos específicos, de esta manera buscando responder preguntas surgidas desde la ciencia de datos [11]. Por su naturaleza multidisciplinaria, incluye una variedad de técnicas con muchos objetivos diferentes donde se identifican 4 áreas principalmente:

- **Analítica descriptiva**, la cual busca mostrar preliminarmente preguntas sobre qué sucedió en determinada área a partir de elementos del análisis y visualización de datos.
- **Analítica de diagnóstico**, rama que permite responder profundamente a por qué sucedieron eventos a partir de los datos con indicadores y pasos más allá que solo la descripción.
- **Analítica predictiva**, área que busca predecir lo que sucederá en el futuro a partir de la información del pasado utilizando técnicas de la estadística y el machine learning.
- **Analítica prescriptiva**, que busca responder a la pregunta de qué se debe hacer actualmente para obtener determinados resultados en el futuro, apalancándose desde los conocimientos de las otras etapas de la analítica y estrategias desde el machine learning.

La analítica de datos se ha vuelto relevante en el sector empresarial los últimos años al usar los conocimientos de los datos para generar decisiones informadas y en tiempo real en distintos procesos. Las áreas más notables de aplicación son las finanzas, marketing, operaciones, entre otros.

2.3. Metodologías para proyectos de ciencia de datos

La complejidad de proyectos basados en técnicas de la ciencia de datos, propia por involucrar procesos de largo aliento en el trabajo con datos, requiere considerar caminos o metodologías que contribuyan a la exitosa realización de proyectos con tecnologías como la inteligencia artificial o el machine learning.

Estas metodologías definen las mejores prácticas para la aplicación de proyectos de la ciencia de datos independiente de la industria donde sean aplicados, vinculando la parte técnica con la necesidad del negocio u organización [12]. Dentro de las más conocidas en el ambiente empresarial se presentan KDD, SEMMA o CRISP-DM, las cuales se describen brevemente a continuación.

- **Descubrimiento de bases de datos**, o KDD de su acrónimo del inglés, consiste en un proceso metodológico de 5 etapas que permite identificar patrones valiosos, novedosos y potencialmente útiles a partir de un gran volumen de datos para generar conocimiento y tomar decisiones [13].

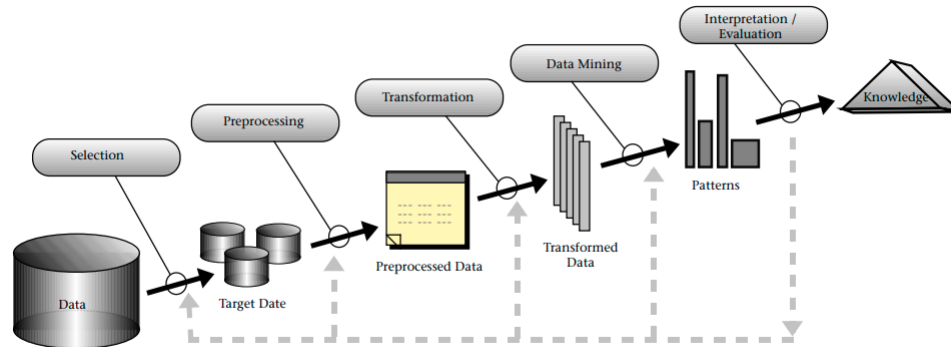


Figura 2.5: Etapas de metodología KDD. [14].

Este descubrimiento de conocimiento parte en primera instancia por comprender la problemática y contexto del proyecto para proceder con la etapa de **selección** y extracción de datos, **preprocesamiento** de la data para su limpieza y validación, **transformación** de los datos, aplicación y prueba de **algoritmos de minería de datos** acordes a los objetivos del proyecto y la **evaluación e interpretación** de los resultados para su uso, visualización y comprensión por el usuario final.

- **SEMMA**, proveniente del acrónimo del inglés “Sample, Explore, Modify and Access”, es otra metodología altamente utilizada que busca detectar patrones novedosos y desconocidos que permitan utilizarse en aplicaciones concretas [15].

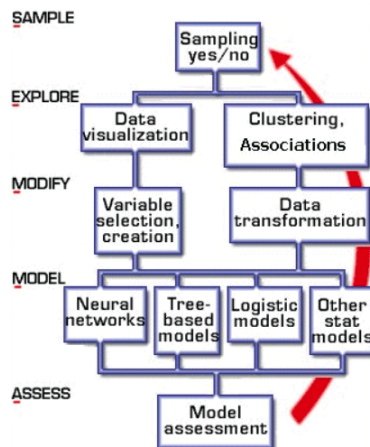


Figura 2.6: Etapas de metodología SEMMA [15].

Este proceso se caracteriza justamente por las etapas de su acrónimo, comenzando por **seleccionar** los datos originales, realizar **análisis exploratorio** sobre estos datos y sus variables, **modificar** la información para su enfoque en los modelos afines al proyecto, **implementar modelos** bajo técnicas analíticas y finalmente **evaluar los resultados** de los modelos midiendo su utilidad y confiabilidad de los objetivos principales.

- **Cross-Industry Standar Process for Data Mining**, o CRISP-DM, es probablemente una de las metodologías más utilizadas en la industria que se caracteriza por tener un origen desde experiencias prácticas de las empresas en el desarrollo de proyectos con algoritmos de machine learning, vinculando fuertemente aspectos de negocio con aspectos técnicos [13][16].

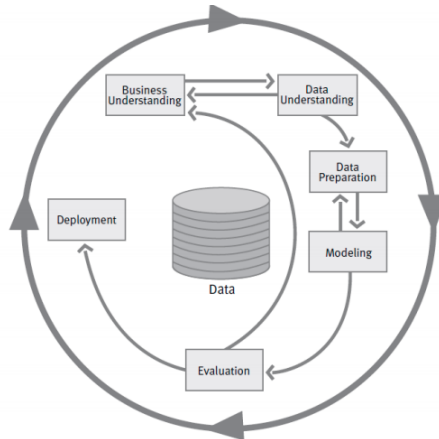


Figura 2.7: Etapas de metodología CRISP-DM [16].

CRISP-DM se caracteriza por poseer 6 etapas de manera general, comenzando por **la comprensión del negocio** y objetivos del proyecto a partir de su contexto y alcances, la recolección de datos y su **entendimiento** mediante un análisis exploratorio, la **preparación de datos** mediante selección, transformaciones, limpieza y selección de variables de los datos iniciales para la **prueba de modelos de machine learning**. Finalmente, se procede con la **evaluación** de los resultados de los modelos técnicamente y la **implementación** de los modelos enfocado en transmitir el conocimiento en acciones concretas en la organización.

2.4. Inteligencia artificial

La inteligencia artificial consiste en un campo que combina la informática, la ingeniería, la estadística, elementos de la filosofía y la psicología, y conjuntos de datos para la resolución de problemas y tareas que usualmente requieren la inteligencia humana [17][18].

Los últimos años este campo de la informática ha tomado una gran atención principalmente por su potencial para la resolución de una gran diversidad de problemas en diversas áreas de las ciencias, esto de la mano del desarrollo de la computación y su capacidad de cómputo, lo que la ha convertido en uno de los booms de las últimas décadas.

Igualmente en relación a sus características, la inteligencia artificial se divide en dos tipos principalmente: inteligencia artificial débil e inteligencia artificial fuerte. La primera, es el tipo de inteligencia que se caracteriza por focalizarse en resolver y aprender tareas específicas con gran precisión y sin ir más allá de las reglas asignadas, con conocidos usos en motores de búsqueda, programas de reconocimiento de imágenes, chatbots o asistentes virtuales.

En cambio la inteligencia artificial fuerte o inteligencia general, consiste en un tipo de inteligencia artificial ideal y de mayor complejidad donde la máquina posee las capacidades cognitivas de planificar, simular y aprender antes de tomar cualquier tipo de decisión de manera autónoma. Este tipo de inteligencia se considera una réplica a la inteligencia humana, sin embargo, aún se mantiene en el campo teórico dado la imposibilidad material de implementarla prácticamente con los recursos computacionales que existen.

2.5. Aprendizaje de máquinas

Aprendizaje de máquinas, o del inglés “Machine Learning”, corresponde a una rama de la inteligencia artificial que busca dotar a máquinas de la capacidad de aprendizaje para la resolución de tareas específicas a partir de un conjunto de experiencias. Este aprendizaje se caracteriza por ser un proceso no guiado donde la máquina aprenda a partir de un conjunto de información de manera autónoma, sin recibir instrucciones específicas.

Dentro del machine learning existen 4 categorías principales a partir de las características del aprendizaje y el volumen de datos involucrados, siendo el aprendizaje supervisado, aprendizaje no supervisado, aprendizaje reforzado y el aprendizaje profundo o deep learning.

Aprendizaje supervisado

Tipo de aprendizaje caracterizado por poseer un conjunto de datos con etiquetas que identifican su salida, y mediante estos, permitir que los modelos logre predecir la salida asociada a nuevas entradas a partir de funciones matemáticas. De esta manera se entiende que el modelo aprende de manera guiada o supervisada.

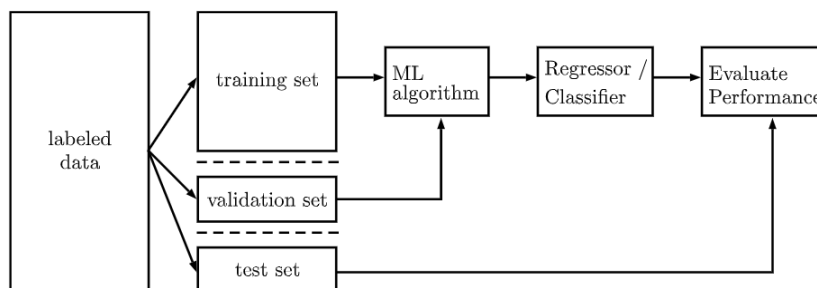


Figura 2.8: Esquema representativo de aprendizaje supervisado [19].

Dentro del aprendizaje supervisado los principales algoritmos de este son las clasificaciones, técnica que permite predecir la clase cualitativa de una entrada y regresiones, que corresponde a una técnica utilizada para predecir, pronosticar y reconocer patrones entre datos cuantitativos.

Aprendizaje no supervisado

Aprendizaje enfocado en algoritmos que clasifican grandes volúmenes de datos sin etiquetas, y por ende sin conocimiento previo de su salida, buscando reconocer patrones y estructuras que permitan organizar estos datos de acuerdo a una función que establezca criterios sobre esta organización [20].

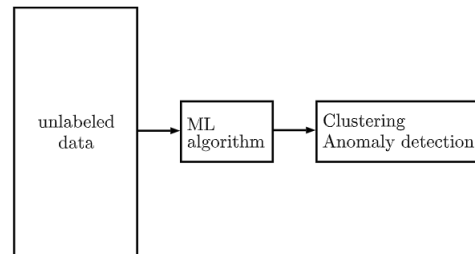


Figura 2.9: Esquema representativo de aprendizaje no supervisado [19].

De los algoritmos más relevantes dentro del aprendizaje no supervisado se encuentran la segmentación o clustering, técnica enfocada en la detección de grupos donde sus miembros posean características similares, y las reglas de asociación, que corresponden a métodos que permiten encontrar relaciones relevantes entre variables de los datos.

Aprendizaje reforzado

Tipo de aprendizaje basado en el proceso de elección de acciones que buscan maximizar una función relacionada a incentivos/castigos de dichas acciones. Este aprendizaje no posee etiquetas asociadas los datos de entrada y tiene la particularidad de interactuar fuertemente con el entorno, permitiendo obtener errores que maximicen los incentivos.

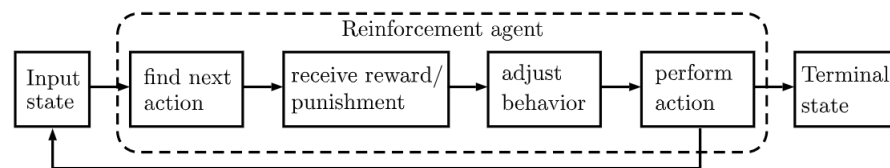


Figura 2.10: Esquema representativo de aprendizaje reforzado [19],

Este tipo de aprendizaje está fuertemente relacionado a los algoritmos de vehículos autónomos y movimiento autónomo de robots, donde la máquina aprende de su entorno la tarea relacionada, en este caso, asociada a un desplazamiento y la retroalimentación de este.

Aprendizaje profundo

Del inglés “Deep Learning, este tipo de aprendizaje se caracteriza por la utilización de redes neuronales artificiales que se basan en el principio de adquisición de conocimientos de las neuronas humanas, permitiendo a la máquina aprender problemas complejos y no estructurados a un mayor nivel que otros algoritmos.

El deep learning funciona bajo un proceso secuencial e iterativo con gran versatilidad de aprendizaje gracias al uso de muchas capas de procesamiento, lo que permite el trabajo con un gran volumen de datos no necesariamente estructurados.

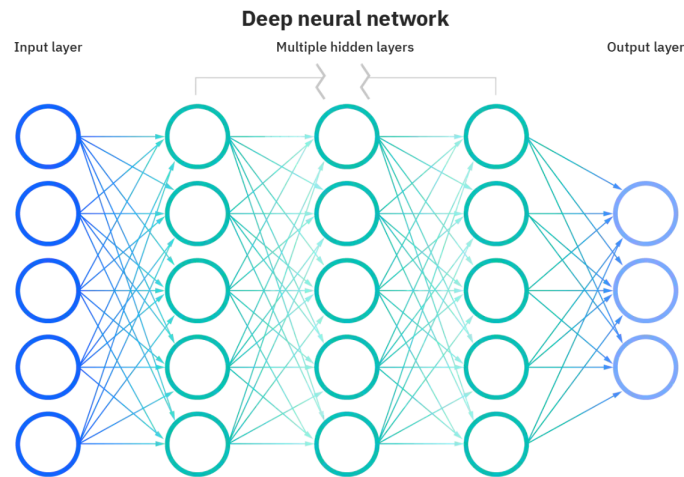


Figura 2.11: Esquema representativo de redes neuronales, base del deep learning [21].

Sus principales usos van en la línea del reconocimiento de imágenes, reconocimiento de voz, sistemas de recomendación, detección de objetos, algoritmos de conducción autónoma, robótica, entre otros.

2.6. Algoritmos de Machine Learning

En la realización de un proyecto de ciencia de datos parte de sus etapas consiste en la implementación, idealmente en mayor cantidad, de modelos basados en machine learning capaces de resolver tareas específicas relacionadas con el proyecto.

La elección de la o las técnicas apropiadas se relaciona directamente con los objetivos del proyecto y el tipo de datos disponibles para su realización. En este caso, considerando el contexto y alcance de los proyectos a implementar enfocados principalmente en trabajo con datos de series de tiempo, se detallan las principales técnicas asociadas a regresiones, patrones secuencias y pronóstico de series de tiempo.

2.6.1. Técnicas de regresión

Los modelos de regresión corresponden a algoritmos de aprendizaje supervisado que se caracterizan por predecir una variable dependiente a partir de una o más variables de entrada independientes. Es una de las áreas de aplicación del aprendizaje de máquinas con mayor uso por la ciencia de datos por su utilidad y versatilidad [22].

Dentro de las regresiones, éstas se diversifican de acuerdo a cantidad de entradas; lineal con una sola variable independiente y múltiple con más de una, al tipo de salida del modelo; como las regresiones logísticas que permiten obtener variables categóricas tales como salidas binarias de verdadero o falso, entre otras.

2.6.1.1. Regresión lineal

La regresión lineal es el tipo de regresión más conocida y base de esta rama de algoritmos la cual se distingue por identificar patrones y relaciones entre variables independientes (x) y una variable continua a predecir (y) de acuerdo a una relación lineal como se expresa en la ecuación 2.1.

$$y = b_0 + \sum_i^N b_i \cdot x_i + e \quad (2.1)$$

Donde b corresponde a los pesos o coeficiente asociados a la predicción, b_0 es la intersección y e al error residual. El objetivo de la regresión lineal consiste en minimizar el error de la predicción del valor cuantitativo en comparación al valor real [23].

La regresión lineal gráficamente se aprecia en la figura 2.12, donde se aprecia la recta de predicción que se forma de acuerdo a las variables independientes y sus pesos óptimos.

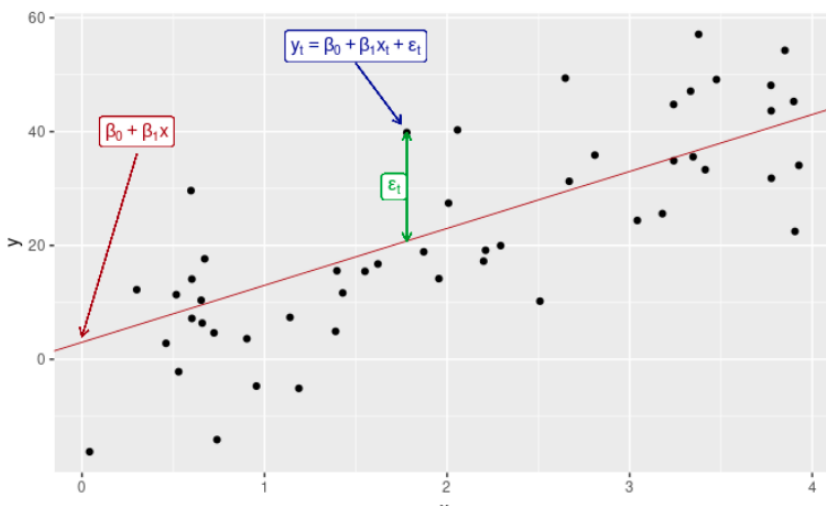


Figura 2.12: Ejemplos de regresión lineal [23].

2.6.2. Pronóstico de series de tiempo

El pronóstico de series de tiempo, o del inglés “timeseries forecasting”, es una rama de algoritmos enfocados en la predicción de valores cuantitativos en secuencias de datos con ocurrencias cronológicas. Esta técnica es de las con mayor uso en diversas áreas de las ciencias, economía y en general de aquellas que datos relacionados temporalmente [10][23].

La base de los modelos de series de tiempo consideran distintas ramas: desde modelos basados en descomposición de la serie de tiempo en su tendencia histórica, estacionalidad y otros, modelos basados en regresiones, modelos de suavización de la curva de acuerdo a su media móvil, modelos de machine learning y finalmente modelo que involucran y unen más de uno de los anteriores, tal como se observa en la figura 2.13.

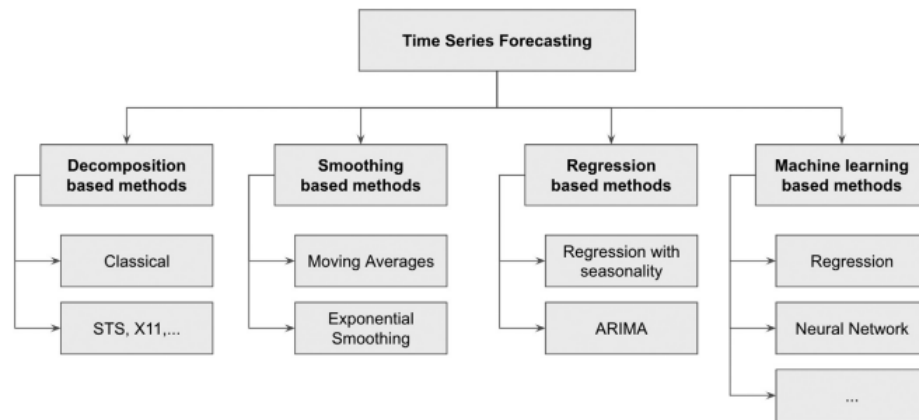


Figura 2.13: Marco referencia modelos de series de tiempo [10].

Específicamente de los modelos basados en regresiones y machine learning, estos buscan inferir relaciones de los datos para predecir valores del futuro a partir de la historia y patrones relevantes de estas observaciones cuantitativas a través del tiempo. Para esto, traducen todos estos inputs en funciones internas que permiten generar la predicción a futuro como se muestra en la figura 2.14.

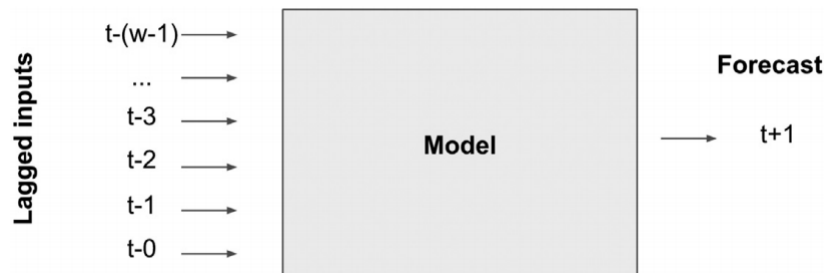


Figura 2.14: Representación modelo de timeseries forecasting[10].

2.6.2.1. Prophet

En el campo de regresiones aplicadas a series de tiempo surge Prophet, modelo de machine learning realizado por Facebook enfocado en la predicción de series de tiempo con estacionalidades y eventos importantes. Prophet se basa en un modelo modular de regresión que permite obtener predicciones precisas y fácilmente ajustables [24].

La base de este modelo de predicción se basa en la descomposición de series de tiempo en componentes de tendencia, estacionalidad, eventos y el error de pronóstico como se expresa en la ecuación 2.2.

$$y(t) = g(t) + s(t) + h(t) + e(t) \quad (2.2)$$

donde:

- $g(t)$ corresponde a la función asociada a la tendencia de cambios no periódicos de la serie de tiempo y que se basa en dos posibles tipos de funciones: logística no lineal por partes o lineal por partes, esto para integrar los cambios de tendencia en el tiempo.
- $s(t)$ es la función relacionada a las estacionalidades (anual, mensual, diaria, etc.) de la serie de tiempo que utiliza de fondo sumas parciales de series de fourier para identificar la tasa de cambios de los ciclos de las estacionalidades.
- $h(t)$ representa la modelación de eventos especiales y feriados en la serie de tiempo, los cuales se caracterizan por ser fechas con mayor dispersión e irregularidad. Estos pueden representarse mediante funciones binarias de verdadero y falso según la fecha.
- $e(t)$ es el error o ruido asociado al modelo de pronóstico la cual recoge alteraciones, incidencias ocasionales o eventos sin explicación periódica ni de tendencia en la serie de tiempo.

Un potencial que posee prophet es la inclusión de regresores extra, los cuales permite incluir al modelo de predicción efectos asociados a datos de otra naturaleza siempre que cumplan con estar relacionados temporalmente a la información base del modelo.

Por último, la adición de regresores extra e incluso la consideración de la estacionalidad se puede realizar de manera multiplicativa, ante lo cual la formulación del modelo de prophet se ve modificado como lo expresa la ecuación 2.3.

$$y(t) = g(t) \cdot \underbrace{(1 + s_m(t) + b \cdot r_m(t))}_{\text{Términos multiplicativos}} + \underbrace{s_a(t) + r_a(t) + h(t) + e(t)}_{\text{Términos aditivos}} \quad (2.3)$$

con $r(t)$ asociado a regresiones exógenas de carácter extra, b corresponde al peso o coeficiente y los subíndices $\{a, m\}$ en $s(t)$ y $r(t)$ se relacionan a los términos aditivos y multiplicativos respectivamente.

2.6.2.2. NeuralProphet

NeuralProphet corresponde a un modelo híbrido de predicción de series de tiempo que se inspira en ser un puente entre modelos clásicos de predicción en base a descomposiciones y modelos basados en deep learning [25].

En particular, NeuralProphet implementa módulos de redes neuronales para series de tiempo mediante Pytorch (librería enfocada en modelos de deep learning) sin perder las características esenciales de Prophet como su interpretabilidad, adaptabilidad, capacidad generar predicciones precisas y a fin de cuenta incorporar las mejores prácticas del machine learning para usuarios no expertos.

Al igual que Prophet, NeuralProphet se basa en la descomposición modular aditiva de la serie de tiempo en tendencia, estacionalidad, eventos y feriados con elementos desde el uso de redes neuronales como autoregresiones y covariables involucradas en la predicción [25]. De esta manera la predicción para valores futuros $y(t)$ se compone de la expresión 2.4.

$$y(t) = T(t) + S(t) + E(t) + F(t) + A(t) + L(t) \quad (2.4)$$

donde:

- $T(t)$ es función relacionada a la tendencia de cambios no periódicos de la serie de tiempo conformada por una función lineal por partes.
- $S(t)$ es la componente de efectos de las estacionalidades en la serie de tiempo en base a suma de términos de fourier de acuerdo a su periodicidad.
- $E(t)$ incorpora los efectos de feriados y eventos en la serie de tiempo, los cuales ocurren de manera esporádica y tienen efectos en la fecha cuando ocurre el evento. Suele asignarse con variable binaria 0 o 1.
- $F(t)$ permite considerar efectos en la regresión de la serie de tiempo para variables exógenas del futuro. Los valores de regresiones futuras deben conocerse (tanto pasadas como futuras) y permite añadirse a las predicciones futuras tanto de manera aditiva como multiplicativa.
- $A(t)$ incorpora el efecto de autoregresión de la serie de tiempo basada en observaciones pasadas. Esta componente se define como $AR(p)$, donde p se ajusta de acuerdo a cuanta historia del pasado afecta los valores del futuro.
- $L(t)$ representa los efectos o correlaciones de regresiones en el tiempo para variables exógenas pasadas, es decir, efectos de covariables como otras series de tiempo que incidan en valores futuros de la variable a predecir. Estas covariables se estructuran de manera similar al modulo $AR(p)$, donde se las p últimas observaciones inciden en la predicción a futuro.

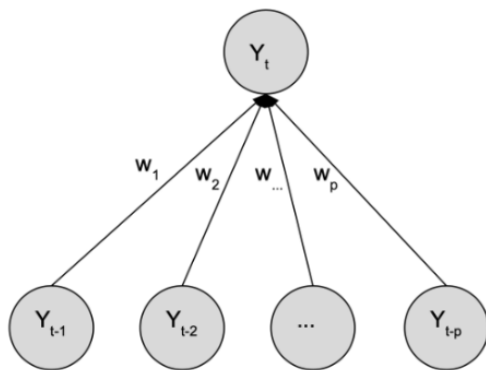
Para la componente autoregresiva $A(t)$ y las covariables $L(t)$ es donde entra el papel de la red neuronal o AR-NET [26]. Esta componente tiene la posibilidad de estructurarse de 3 maneras:

- Como un modulo autoregresivo clásico de manera lineal, donde la red neuronal no posee capas ocultas y funciona como una red con p entradas (historia pasada) y genera h salidas de valores futuros, tal como se expresa en la figura 2.15.a.
- Un modulo con capas ocultas activas en la red neuronal, la cual permite modelar dinámicas no lineales de acuerdo a la configuración propia de la AR-NET como el número de capas, las dimensiones ocultas, el sesgo, el papel de la función de activación (la cual es una función ReLU), entre otras.

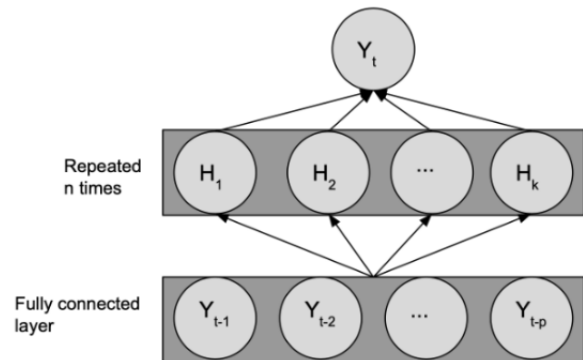
El trade off de la mayor precision de añadir capas ocultas se paga en la interpretabilidad del modelo, donde determinar el aporte de cada componente del pasado se pierde entre las conexiones de las capas y la red neuronal.

- Un modulo de AR-NET donde mediante la utilización de una regularización de los pesos de las conexiones, se permite aproximar el orden de la autoregresión a un valor ligeramente mayor pero que permite que el modelo sea conveniente en tiempo y conserve de mejor manera su interpretabilidad.

Un ejemplo gráfico del funcionamiento de la AR-NET se visualiza en la figura 2.15, donde se observa una estructura con y sin capas ocultas en la red, observando la dinámica de conexiones para p historia del pasado.



(a) Ejemplo de arquitectura de AR-NET con pesos w_i para sus conexiones.



(b) Ejemplo AR-NET con n capas de tamaño k

Figura 2.15: Ejemplo de funcionamiento de AR-NET [26].

2.7. Validación y métricas de error

Parte del trabajo con proyectos de ciencia de datos consiste en la evaluación de los algoritmos de machine learning implementados, buscando escoger el óptimo de estos por su precisión en la tarea objetivo a desempeñar. Para esta evaluación existen diversas métricas de acuerdo al tipo de algoritmo, los datos utilizados y aspectos asociados al contexto del problema.

En los modelos de regresión y predicciones de series de tiempo, la evaluación se enfoca en la precisión o error de pronóstico entre valores predichos \hat{y} y el valor real y tal como se indica en la ecuación 2.5. Esta medida corresponde a una magnitud que sirve de base a otras métricas de error [22][27].

$$e_i = \hat{y}_i - y_i \quad (2.5)$$

Para la medición de la dirección del error se utiliza el sesgo, o del inglés bias, el cual representa el error promedio histórico de las predicciones como se indica en la ecuación 2.6.

$$bias = \frac{1}{n} \sum_{i=1}^n e_i \quad (2.6)$$

donde n corresponde a la cantidad de variables asociadas a la medición de error.

A continuación, se expresan las 3 principales métricas de evaluación MAPE, MAE y RMSE, de las cuales su elección óptima dependerá del caso de estudio y naturaleza del problema.

MAPE

El error porcentual absoluto medio, o MAPE del acrónimo en inglés, corresponde a la medida promedio del error absoluto normalizado por el valor real para cada periodo i de n variables asociadas al error.

$$MAPE = \frac{1}{n} \sum_i^n \frac{|e_i|}{y_i} \quad (2.7)$$

Esta métrica es una de las más utilizadas en la industria, sin embargo, por su formulación matemática de la ecuación 2.7, se observa que esta métrica se sesga frente a la magnitud del valor real y , donde errores grandes en periodos con bajo valor real adquirirán mayor peso.

MAE

El error absoluto medio, o MAE del acrónimo en inglés, es una métrica de evaluación que analiza la media del error absoluto, es decir, las distancias absolutas entre la predicción realizada y el valor real.

$$MAE = \frac{1}{n} \sum_i^n |e_i| \quad (2.8)$$

Esta métrica permite analizar directamente la precisión absoluta del modelo, donde su magnitud esta en unidades de los valores originales y predichos permitiendo relacionar el error con análisis cuantitativos asociados a la naturaleza de los datos.

RMSE

La raíz del error cuadrático medio, o RMSE de su acrónimo del inglés consiste en la raíz cuadrada del error cuadrático medio. Al igual que el MAE, no se normaliza por el valor real lo cual permite que su resultado este en unidades de las variables reales y sea fácilmente comparable.

$$RMSE = \sqrt{\frac{1}{n} \sum_i^n e_i^2} \quad (2.9)$$

Un aspecto relevante del RMSE que se desprende de la ecuación 2.9, es que al elevar al cuadrado el error en la sumatoria esta métrica penaliza con mayor fuerza los errores altos de la predicción en su valor final, siendo una métrica con gran valor cuando se busca evitar errores considerables.

Capítulo 3

Metodología

3.1. Propuesta metodológica

Para la metodología de esta memoria se seguirá el procedimiento en la figura 3.1, inspirada en las metodologías para proyectos con datos y algoritmos de machine learning vistos en la sección 2.3.

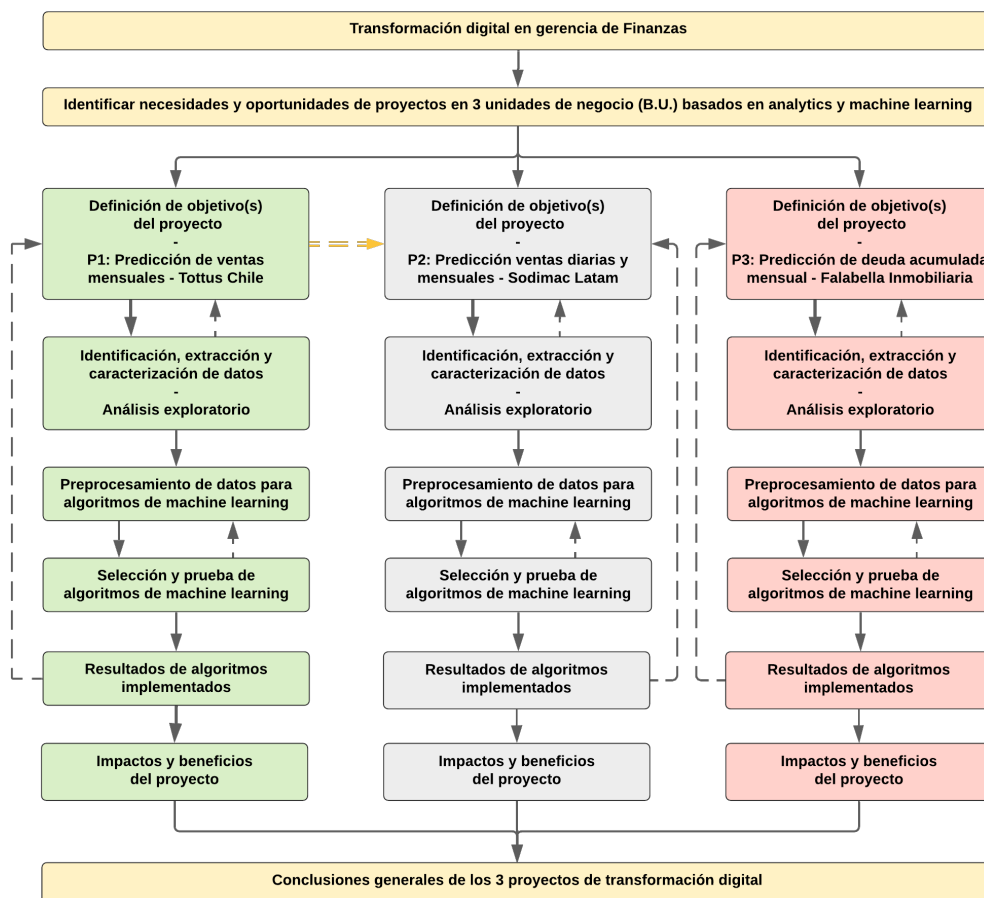


Figura 3.1: Diagrama de metodología propuesta a seguir.

En el contexto de la transformación digital corporativa de Falabella y alineado con los objetivos del presente trabajo se comienza por identificar necesidades y/o oportunidades en procesos financieros de 3 unidades de negocio distintas, lo que permita el diseño e implementación de los correspondientes 3 proyectos de transformación digital.

Definición de proyectos y objetivos

Para cada uno de estos proyectos, la primera etapa consiste en identificar y caracterizar correctamente la problemática y objetivos del proyecto, vinculando estos con una solución a partir de algoritmos de machine learning y analítica avanzada.

Junto a esta definición, se revisa el contexto del proyecto caracterizando el proceso a impactar, tareas relacionados y recursos involucrados buscando tener una visión general que permita medir los impactos de la solución y su relación con la adopción de tecnologías en el proceso involucrado.

Identificación, extracción y caracterización de los datos - Análisis exploratorio

Definidos los objetos del proyecto, se procede con la parte enfocada en el trabajo con datos. En primer lugar se hace un rastreo para la recolección de los datos desde las fuentes de la unidad de negocio del proyecto, ya sean fuentes de bases de datos clásicas como Oracle, SAS, Hyperion o fuentes nuevas en servicios de la nube como bigquery de Google Cloud Platform (GCP).

Junto a su identificación y extracción, se realiza una primera revisión de los datos desde sus atributos, cantidad y calidad de los mismos revisando presencia de outliers, valores nulos y/o faltantes.

Por último, se realiza un análisis exploratorio de los datos identificando frecuencias de los datos, patrones básicos relevantes, análisis estadístico y visualizaciones que permitan dar una mirada hacia la solución a partir del uso de algoritmos de machine learning.

Preprocesamiento de datos para algoritmos de machine learning

Tras la caracterización general de los datos, se realizan técnicas y pasos enfocados en preparar los datos para los algoritmos de machine learning. Estas modificaciones dependen del tipo de algoritmo a desarrollar, siendo las frecuentemente realizadas:

- **Selección de subconjuntos de datos y variables**, útiles y eliminando las que no agreguen valor para los algoritmos.
- **Limpieza de datos**, donde se enfoca en la normalización de los datos, discretización de atributos, tratamiento de valores nulos u outliers, y elementalmente toda técnica que permita preparar el formato de los datos para los requerimientos de los modelos.
- **Formateo de los datos**, generando transformaciones en los datos sin modificar su significado, ya sea ordenando atributos o eliminando caracteres del lenguaje como comas, tildes, espacios, etc. que pueden generar problemas en la implementación de modelos.

Selección y prueba de algoritmos de machine learning

A partir de los datos del preprocesamiento realizado, se procede a la selección y prueba de algoritmos de machine learning en cada uno de los proyectos, enfocados en el cumplimiento de los objetivos de cada proyecto

Los posibles algoritmos a modelar, tal cual se hacía referencia en los alcances del trabajo, se enfocan en algoritmos de pronóstico de series de tiempo cuyo objetivo es predecir valores cuantitativos en relación a datos indexados con fechas temporales.

En concreto, el pool de algoritmos utilizados en este trabajo se desglosa en:

- Prophet: Modelo basado en descomposición de series de tiempo y modelos aditivos generalizados (GAM).
- NeuralProphet: Modelo basado en Prophet (GAM) con la mejora de poder utilizar una red neuronal autoregresiva (AR-NET) para componentes de la predicción.

Pese a existir otros tipos de algoritmos para pronóstico de series de tiempo como redes neuronales recurrentes de corta memoria (RNN-LSTM), arboles de decisión o modelos clásicos de descomposición como ARIMA y SARIMAX, se argumenta la elección de algoritmos en base a ser modelos modernos, flexibles en su ajuste de parámetros y cumplir principalmente con la característica de ser algoritmos interpretables que no funcionan como caja negra para usuarios no expertos.

Por consiguiente, se procede en cada uno de los proyectos se sustenta la selección del o los algoritmos a implementar con su respectivo ajuste de parámetros y la evaluación de los resultados en base a métricas de error, seleccionando en cada caso los mejores resultados de predicción de series de tiempo a futuro.

Resultados de algoritmos implementados, impactos y beneficios de los proyectos

En base al los modelos de machine learning desarrollados, se procede con la validación de sus resultados en base a la necesidad inicial de la unidad de negocio, procesos involucrados, métricas de error y aspectos propios del negocio.

De manera general esta evaluación tiene dos grandes niveles: por un lado se mide los resultados de los modelos bajo métricas relacionadas al pronóstico de series de tiempo: error, sesgo, error absoluto medio (MAE), etc. en comparación a procesos actuales en la unidad de negocio (en caso de existir) y entre los algoritmos de machine learning implementados (cuando se implemente más de uno).

Posteriormente se analizan los resultados en un nivel de impactos y beneficios que conlleva su desarrollo: analizando aspectos de mejoras de los procesos involucrados como reducción de horas hombre, mejoras de precisión de las predicciones y elementos propios de la transformación digital como la automatización de tareas, confiabilidad y rapidez gracias al uso de procesamiento en computadoras y servicios de la nube, entre otros.

Conclusiones generales de los 3 proyectos de transformación digital

Finalmente, se analiza globalmente los resultados e impactos de los proyectos en las gerencias de finanzas de las unidades de negocio de Falabella S.A., considerando alcances de los proyectos, sus procesos involucrados, metodologías desarrolladas, y en general todo aspecto relacionado a la transformación digital de la gerencia mediante la incorporación de tecnologías y la toma de decisiones en base a datos.

3.2. Implementación

A partir de la metodología desarrollada, se procede con su implementación en cada uno de los proyecto hasta la selección y prueba de algoritmos de machine learning, dejando para el siguiente capítulo los resultados, evaluación de los mismos e implicancias en los procesos asociados a las unidades de negocio.

3.2.1. Identificación de necesidades y oportunidades de proyectos en 3 unidades de negocios

Para la definición de los 3 proyectos de transformación digital, se diagnostican oportunidades y/o necesidades en 3 unidades de negocios donde existan procesos que puedan ser mejorados a partir de soluciones desde la analítica avanzada y el machine learning.

P1: Pronóstico de ventas para Tottus Chile

El primer proceso detectado, propio en áreas financieras, fue el de pronóstico de variables financieras en negocios con características de venta minorista como lo es Tottus Chile. El pronóstico de ventas, margen o costos corresponde a un proceso rutinario que se utiliza para tareas como generar presupuestos de áreas, seguimiento de planes de venta, tareas de control de gestión como analizar el comportamiento de venta del pasado y tomar acciones comerciales a futuro, entre otros.

A fin de cuentas, en Tottus Chile se encontraban con la oportunidad de modernizar su proceso de pronóstico de ventas mensuales para sus tiendas, un proceso que actualmente se realizaba en excel con reglas de negocio basadas en decisiones y conocimiento experto.

P2: Pronóstico de ventas para Sodimac

A partir de los resultados y caso de éxito del modelo para ventas implementado en Tottus, se genera el interés en Sodimac de replicar un modelo predictivo de ventas diarias y mensuales.

Pese a tener similares objetivos que el modelo en Tottus, el contexto de Sodimac poseía la particularidad de no poseer actualmente ninguna metodología para pronosticar ventas en los canales de venta de Sodimac, esto por ser un proceso engorroso sin capacidad de horas hombres disponibles para dedicarle,

P3: Pronóstico de deuda acumulada real para Falabella Inmobiliaria

Finalmente para el tercer proyecto abarcado en la memoria, este nace a partir de las intenciones de Falabella Inmobiliaria de adoptar en sus procesos, o parte de ellos, elementos desde tecnologías digitales ya sea automatizando procesos, generando visualizaciones y/o aplicaciones desde el machine learning, haciendo uso provechoso de los datos de la unidad de negocio.

En concreto, se observa la oportunidad de generar modelos predictivos para la deuda acumulada del negocio, mejorando este proceso en sus distintas etapas de trabajo con datos a partir del uso de tecnologías digitales.

3.2.2. Desarrollo de proyectos

3.2.2.1. Proyecto: Predicción de ventas de Tottus

Definición de objetivos

Para el proyecto en Tottus se diagnosticó la necesidad implementar una metodología para el proceso de predicción de ventas a partir de la data histórica y algoritmos desde el machine learning, generando un proceso basado en datos que permitiese generar pronósticos objetivos y cuantificables sin depender de reglas de negocio que pueden funcionar o fallar de la misma manera.

De esta manera, el objetivo principal de proyecto es diseñar e implementar un modelo de predicción de ventas netas mensuales para Tottus Chile, con potencial apertura por días y locales basado en datos históricos del negocio, elementos desde la analítica y un algoritmo de machine learning con variables interpretables del negocio.

Identificación, extracción y caracterización de datos - Análisis exploratorio

La base de datos necesaria para el modelo de predicción consiste en la información transaccional de Tottus Chile, la cual contiene la datos de ventas, ventas sin impuesto, costos y plan de ventas de la unidad de negocio a nivel diario junto a otras variables de interés.

Estos datos se obtienen en primera instancia a partir de 3 fuentes en forma de archivos Excel, sin embargo, posteriormente se actualiza la fuente de extracción desde la base de datos en la plataforma de Google Cloud Plataform (GCP) - “Big Query”, la cual incrementa el detalle y variables en la información. A fin de cuentas se posee:

- Información transaccional diaria desde enero 2019, con variables cuantitativas del negocio (venta, venta neta, costos, unidades vendidas, etc.), fechas como datetime, códigos de locales y códigos de subcategorías de productos.
- Información de productos (en formato de texto) identificando su descripción, marca, categorías y distintos niveles de subcategorías por código y nombre real.

- Información relevante a clientes y sus características como edad (cuantitativa), segmento (texto), y otros sin identificar clientes individual y directamente.
- Información transaccional relevante al pago, su método de pago (tarjeta, efectivo, etc.) e identificadores de productos y servicios de la compañía, como elementos de CMR Falabella.
- Información de locales con detalle por código, nombre real y ubicación mediante coordenadas geográficas.
- Información de asignación de costos de acuerdo a productos vendidos, con desglose por sus distintas componentes.
- Información de de envío y conexión con centros logísticos, identificando tipo de envío, información geográfica, costos, entre otros.

A fin de cuentas, la base final en GCP posee 89 variables, de las cuales, la información diaria a nivel transaccional resulta ser la información relevante para la implementación del modelo con vías de pronosticar las ventas netas a nivel mensual.

Con respecto a la calidad de los datos, se observa que existe una gran cantidad de variables con valores nulos pero que no resultan ser variables de interés y podrían eliminarse del trabajo. De igual manera, para la variable de las ventas netas diarias no existen vacíos ni saltos en la misma base más allá de los días de cierre de locales de Tottus por feriados irrenunciables, lo cual es un comportamiento normal explicado por la dinámica de funcionamiento de los supermercados.

De manera general, al analizar las ventas netas de Tottus se aprecia que se posee información desde enero de 2019 a julio de 2021, siendo 929 datos con resolución diaria de los cuales se puede revisar variables estadísticas en la tabla 3.1, con un rango desde el valor nulo de ventas (probablemente para días feriados irrenunciables) a un peak de 5615 millones de pesos de ventas para un día en específico.

Tabla 3.1: Variables estadísticas de las ventas netas

Variable	Venta Neta (millones de pesos)
Promedio	2263.74
Desviación Estándar	693.40
Mínimo	0.0
Percentil 25 %	1753.78
Percentil 50 %	2166.59
Percentil 75 %	2630.01
Máximo	5615.34

En la búsqueda de mayor detalle y patrones relevantes sobre la información, se realiza un análisis exploratorio enfocado en las ventas netas a resolución diaria y mensual, requiriendo en esta última cambiar la distribución de los datos de días a meses.

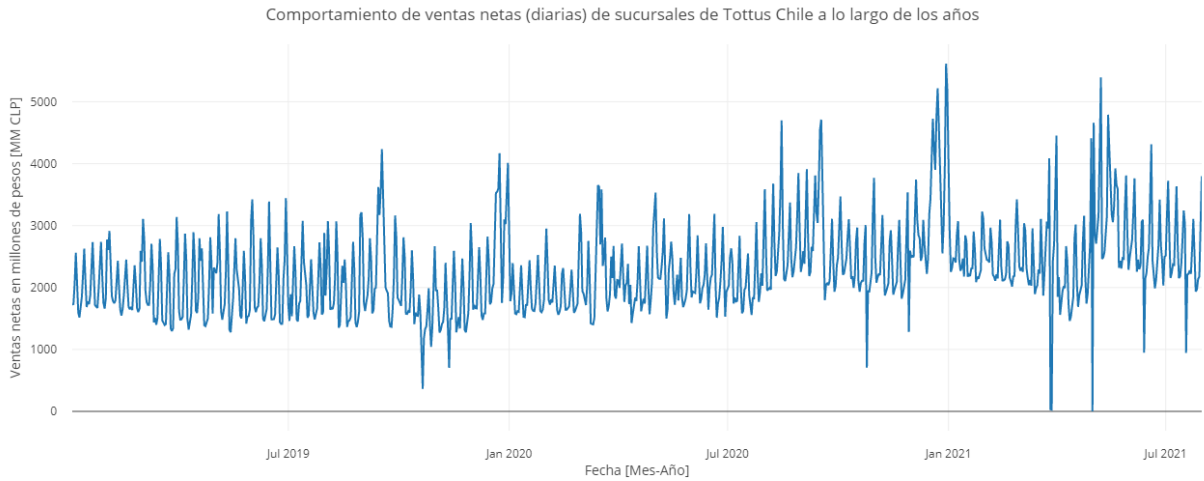


Figura 3.2: Ventas diarias Tottus Chile

De la venta diaria se logra apreciar un comportamiento normal en el tiempo, con disminuciones de ventas bajo los 1000 millones en menos de 10 días para más de dos años de información, principalmente justificado por fechas cercanas a feriados, votaciones o fenómenos sociales donde se desplaza demanda a días contiguos.

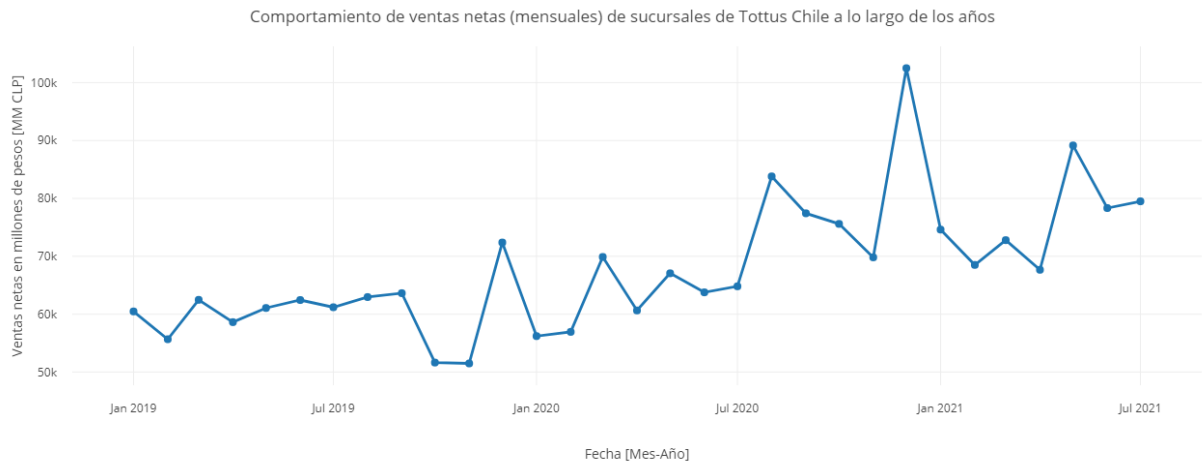


Figura 3.3: Ventas mensuales Tottus Chile

Para la resolución mensual de las ventas, ilustrada en la figura 3.3, se observa una tendencia al alza a lo largo de los años pasando desde los 60.000 a los 80.000 millones de pesos en ventas para un mismo mes. Además, se detectan elementos estacionales para determinados meses del año como aumentos de ventas en los meses de diciembre o marzo, los cuales se explican por compras de navidad o compras escolares de inicio del año escolar.

Otro elemento relevante que se logra observar de ambas figuras, se destaca en incrementos de ventas netas en fechas que no poseen características comerciales en el último año como lo son los meses de agosto del 2020 y mayo del 2021. Estos dos periodos sumados a diciembre del 2020 resultan un patrón que se logra explicar por ser fechas contiguas a los retiros del

10% de AFP en el contexto de la pandemia, que aportaron flujos de dinero en la población, y en cierto medida, incrementos de consumo que se reflejan en estos incrementos inusuales de las ventas netas de los supermercados Tottus.

Los comportamientos de la tendencia al alza y estacionalidad de las ventas netas se ven reflejados en otras variables financieras del negocio como los costos y unidades vendidas, visualizables en las figuras 3.4 y 3.5, validando la idea que se posee una base de información suficiente y que el comportamiento para patrones como feriados, fechas de votaciones o retiros de las AFP es consistente en todas las variables transaccionales.

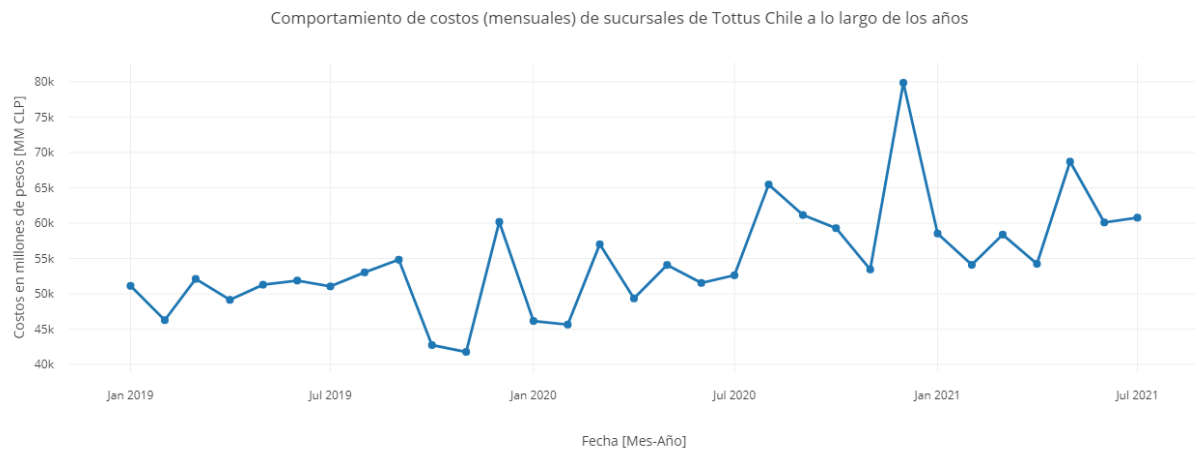


Figura 3.4: Costos mensuales Tottus Chile

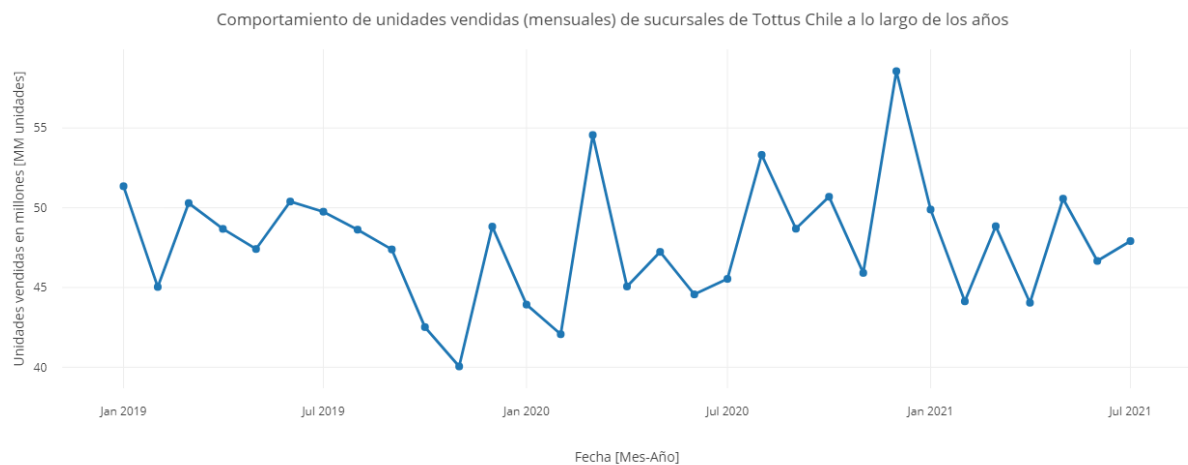


Figura 3.5: Unidades vendidas mensuales Tottus Chile

Preprocesamiento de datos para algoritmos de machine learning

Con el objetivo de generar las proyecciones para las ventas netas de Tottus Chile, se comprende que de las 89 variables iniciales gran parte de estas resultan ser información irrelevante para este proceso.

Se selecciona la variable financiera de la venta neta, la fecha de dichas ventas y los locales o supermercados como datos útiles para el trabajo. Esta última variable se selecciona debido que la cantidad de locales operativos en el país no es un número fijo, sino que existen tiendas que cerraron en el pasado y su consideración agregaría un ruido a los modelos de proyección puesto que habrían montos de ventas de las cuales se tiene historia pero no deberían considerarse en el futuro por su cierre.

De esta manera de los 74 locales de los cuales se posee data disponible desde 2019, se eliminan 5 locales cerrados: Calama Mall, Llolleo, Pedro Fontova, Peñalolen y Puente Alto II, quedando 69 supermercados operativos.

Además, a partir de que no existen valores anómalos por problemas de data ni valores nulos que no se deban a fechas de feriados, no se requiere realizar un proceso de imputación de la data.

Finalmente, para la implementación del algoritmos de machine learning, se elimina la variable de locales físicos de la base de datos final quedándose con la venta neta a nivel diario para todo Tottus Chile. Estos datos se dejan en resolución diaria para posteriormente agruparla en resolución mensual post ejecutar el modelo de predicción, buscando captar correctamente efectos en fechas puntuales como los retiros de AFP y los feriados del año.

Otro aspecto de los datos, es que en vías a la implementación del algoritmo de predicción que será Facebook Prophet, este requiere formatear la data con la estructura de **ds**: **fechas** e **y**: **valores de venta neta** como se refleja en la tabla 3.2.

Tabla 3.2: Ejemplo de estructura de data vías a modelo Prophet

ds	y
01-01-2019	xxx
01-01-2019	xxx
...	...
30-07-2021	xxx
31-07-2021	xxx

Junto a esto, se entiende que un elemento común en la prueba de algoritmos de machine learning consiste en separar los datos en, al menos, dos subconjuntos: uno para el entrenamiento del modelo denominado *train-set* y un segundo subconjunto de datos denominado *test-set* para probar el desempeño del algoritmo en base a métricas de error como es el caso con modelos de pronóstico de series de tiempo.

Se define comparar el desempeño de los modelos a 3 meses, por lo cual se separan los datos con la información desde enero de 2019 a abril de 2021 como data de entrenamiento y los meses de mayo-junio-julio de 2021 como conjunto de prueba para el algoritmo de machine learning a implementar.

Selección y prueba de algoritmos de machine learning

A partir del preprocesamiento realizado, se procede a construir el modelo de predicción de ventas mensuales para Tottus. Para este modelo se decide utilizar el algoritmo de Prophet en el entorno de Python.

Este algoritmo, descrito en detalle en la sección 2.6.2.1, se caracteriza por ser un algoritmo con una curva de aprendizaje amigable para usuarios no expertos, generando predicciones rápidas, robustas, fácilmente ampliables y en general poseer parámetros interpretables por el conocimiento de usuarios no expertos en machine learning pero con conocimientos propios del negocio.

Estas características lo destacan como un algoritmo interesante para implementar en Tottus Chile dado que permite conectar las predicciones con las componentes esenciales de las ventas netas como su tendencia a lo largo de los años, estacionalidades, efectos de feriados, fechas comerciales y elementos exógenos como los retiros de AFP, funcionando como un modelo que no funciona como caja negra y pueda ser comprendido por los usuarios de la unidad de negocio.

Este modelo posee una serie de parámetros que se pueden ajustar para el ajuste de las predicciones a futuro. De estas variables, se destacan 5 de las más relevantes:

- **changepoint_prior_scale**: Parámetro que ajusta la tasa de cambio de la tendencia a lo largo del tiempo. Tras pruebas del parámetro y validación cruzada en el conjunto de entrenamiento se define este parámetro en 0.01.
- **seasonality_mode**: Parámetro que ajusta los cambios estacionales en la serie de tiempo. Se asignó el modo multiplicativo debido a la magnitud de los cambios acorde anuales, semanales y la magnitud de los mismos.
- **seasonality_prior_scale**: Parámetro que controla la flexibilidad de la estacionalidad y su cambio, a mayor valor significa que la estacionalidad se ajusta a grandes variaciones pero puede sobre ajustarse a estos cambios. El parámetro por defecto es 10, sin embargo, se modifica a un valor de 0.1 tras realizar validaciones cruzadas en los datos.
- **holidays_prior_scale**: Parámetro que controla el ajuste e impacto de los feriados en el modelo, similar a `seasonality_prior_scale`. Se define en 0.01 tras la prueba de parámetros.
- **x_seasonality**: Este parámetro, que hace referencia a las estacionalidades diarias, semanales y/o anuales, ajusta la existencia de estas estacionalidades en los datos. Se setean en `true` las estacionalidades anuales y semanales dado que hacen sentido con los cambios y comportamiento en los datos.

Por último, otro aspecto que permite el modelo de Facebook Prophet es la capacidad de integrar regresores extra en el entrenamiento del modelo, tanto para valores numéricos como booleanos. En este aspecto, se introduce un regresor de los feriados legales del país mediante el parámetro de `add_country_holidays` y un regresor denominado *eventos* que integra las

fechas de votaciones, fechas de retiros, etc. indicados como un 1 y 0 para las fechas que no coincidan en el calendario.

3.2.2.2. Proyecto: Predicción de ventas diarias y mensuales en Sodimac Latam

Definición de objetivos

El proyecto de predicción de ventas por países y canales de Sodimac Latam se origina por un lado, observando los resultados e impacto del modelo de predicción de venta en Tottus Chile, y más relevante, por la necesidad de implementar un modelo de pronóstico de ventas en un área que actualmente no realiza esta tarea o proceso de manera constante y actualizada.

Se define como objetivo principal el desarrollo de un modelo de predicción de ventas a nivel país en Sodimac con 2 niveles: por un lado generar predicciones diarias y extender estas predicciones a nivel mensual.

Identificación, extracción y caracterización de datos - Análisis exploratorio

La información principal para el modelo de predicción se diagnóstica como la data consolidada de ventas para todos los países y canales de Sodimac. Estas se presentan para los 7 países y canales de venta que se detallan a continuación:

- Países con presencia: Chile, Argentina, Uruguay, Brasil, Perú, Colombia y México.
- Canales de venta: Empresa, Store Retail, DVD o canales digitales, Kiosko y otros sin clasificar.

Esta base de datos disponible desde bigquery en GCP, presenta 71 variables con data en 5 grandes grupos:

- Información transaccional de ventas, costos, impuestos y margen a nivel diario.
- Información de canales de venta a distintos niveles de apertura.
- Información del cliente y tipo de pago, sin identificar al consumidor con sus datos personales.
- Información de seguimiento de transacciones con inventario, despacho, entre otros.
- Información de país de la venta y locales dentro del mismo país en formato de códigos.

Al revisar la base de datos de GCP se diagnostican presencia de outliers, valores nulos para canales de venta en días sin patrones definidos y existencia de variables vacías. Esto se produce dado que la base en bigquery se encuentra actualmente en formación desde sus fuentes de origen y se encuentra en un proceso de consolidación y validación de la información aún no finalizado.

Se trabaja finalmente con la información consolidada de ventas diarias por país y canales de venta generada en formato Excel, donde se corrigen valores nulos y outliers desde las fuentes de origen disponiendo las ventas diarias consolidadas a nivel de país y subcategoría de canales de venta para Sodimac Latam desde enero 2019 a octubre de 2021 en todos los países.

Analizando los datos de ventas de Sodimac en los distintos países, se observa en primera instancia el comportamiento estadístico para cada país, entendiendo que la cantidad de locales y participación de mercado no es la misma en los 7 países. De esta manera, se logra apreciar como los niveles de ventas de Colombia y Chile, del orden de 11.000 y 6.000 millones, son notoriamente mayores a los otros 5 países donde Argentina esta en torno a 22 millones y ya menor a los 10 millones para Uruguay, Perú, México y Brasil.

Tabla 3.3: Revisión estadística de ventas en países de Sodimac

Variable	Argentina	Brasil	Chile	Colombia	México	Perú	Uruguay
# Datos	1017.0	1017.0	1017.0	1016.0	1017.0	1017.0	1009.0
Promedio	22.69 M	3.68 Mill	6047.05 Mill	11569.14 Mill	5.07 Mill	8.68 Mill	6.75 Mill
Desv. Estandar	10.36 Mill	1.23 Mill	2079.50 Mill	3397.88 Mill	1.81 Mill	3.44 Mill	2.21 Mill
Min	2626.44	0.0	25.64 Mill	0.0	0.0	-3744.82	116531.18
P 25 %	15.75 Mill	2.92 Mill	4817.78 Mill	10028.65 Mill	3.67 Mill	8.22 Mill	5.15 Mill
P 50 %	20.77 Mill	3.66 Mill	5703.15 Mill	11525.31 Mill	5.14 Mill	9.53 Mill	6.25 Mill
P 75 %	27.83 Mill	4.63 Mill	7611.27	13678.16 Mill	6.20 Mill	10.77 Mill	8.10 Mill
Max	71.96 Mill	7.25 Mill	18033.81 Mill	38735.28 Mill	15.10 Mill	16.56 Mill	14.10 Mill

Además, debe tenerse en cuenta el distinto tipo de divisa para cada país, donde un peso chileno es cercano a los 5 pesos colombianos, lo cual sumado a diferencias innatas en fechas de feriados legales del país destaca la presencia de diferencias innatas para las ventas de cada uno de los países.

A partir de las series de tiempo en el análisis exploratorio, se permite validar los patrones de las ventas vistos recientemente para los países como se observa en la figura 3.6, donde la magnitud de las ventas en Chile y Colombia no logra percibir el resto.

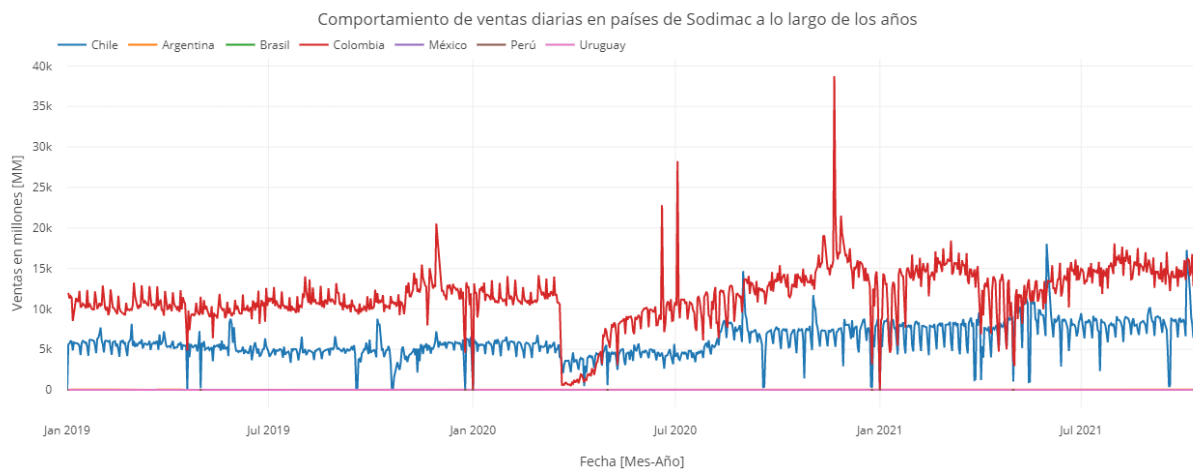


Figura 3.6: Ventas diarias Sodimac Latinoamérica

Filtrando estos dos países se logra analizar las ventas para Argentina, Brasil, Perú, México y Uruguay, claramente menores en magnitud y con diversidad de comportamiento estacional.

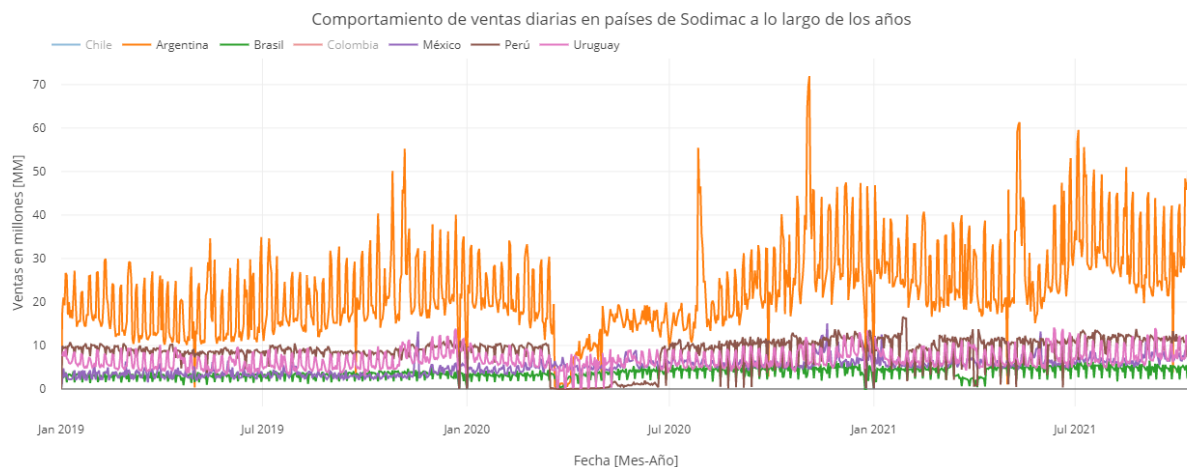


Figura 3.7: Ventas diarias Sodimac AR-BR-PE-MX-UY

Sacando a Argentina de las comparación, se logra hacer mayor zoom al comportamiento de ventas de los últimos 4 países, los cuales presentan ventas menores a 10 millones de sus divisas internas.

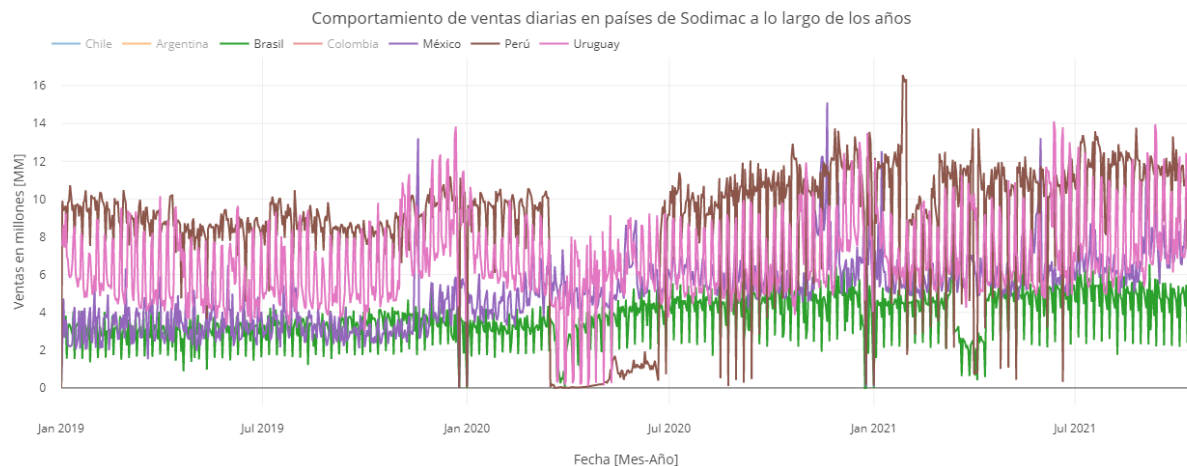


Figura 3.8: Ventas diarias Sodimac BR-PE-MX-UY

Analizando los comportamientos aislados de ventas de cada país, se observan países como Brasil, Uruguay o Perú que poseen mayor dispersión de su curva, propia de estacionalidad semanal con diferencias marcadas entre sus 7 días. A su vez, se ve una tendencias de crecimiento más marcada en países como Brasil y México.

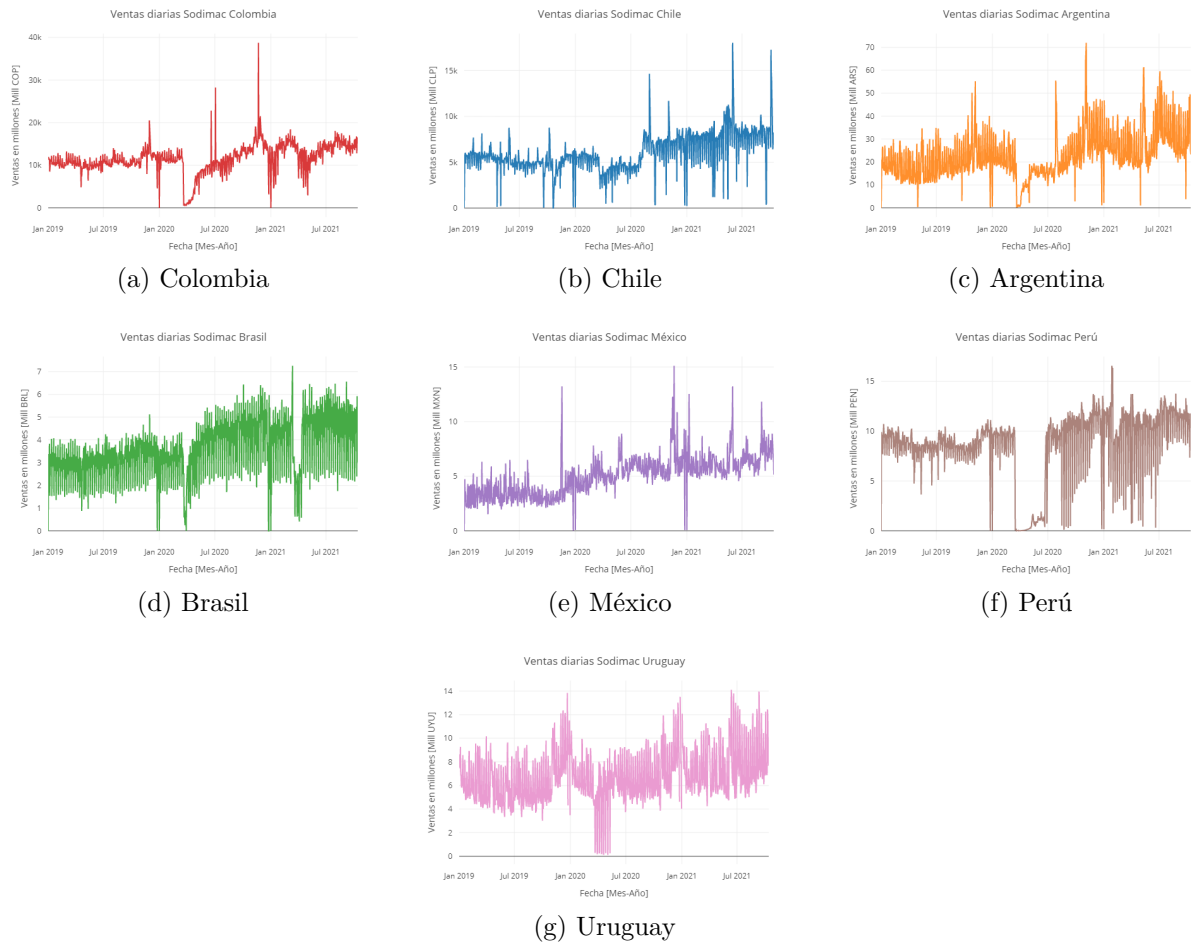


Figura 3.9: Ventas diarias por país - Sodimac.

A grandes rasgos, se observa que las diferencias entre las ventas de países se reflejan en las tendencias a lo largo del tiempo, los cambios estacionales y elementos propios de las series de tiempo.

Pese a tener todos los países tendencias al alza en general a lo largo de los años, no son el mismo nivel de incremento en todos los países. Así mismo, se logra apreciar en países como Argentina, Colombia y Perú un impacto mayor de las cuarentenas por la pandemia del COVID, con una disminución considerable en los primeros trimestres del 2020, impactando su tendencia a futuro con repuntes de distinta magnitud.

Por último, se observa que elementos propios de los países tienen impactos en las ventas; estabilidad de los mercados, clima, retiros de AFP o fenómenos sociales como los ocurridos en Chile y Colombia caracterizan propiamente su comportamiento. Estas diferencias innatas para las ventas de cada uno de los países indica que el ajuste de los algoritmos y sus parámetros debe adecuarse particularmente a cada uno de los países y no generalizar el modelo para todos ellos.

Preprocesamiento de datos para algoritmos de machine learning

La selección y preparación de los datos se enfoca en dos aristas, en primer lugar se define utilizar la base de ventas consolidada de Sodimac sin valores anómalos, nulos o mal imputados además de filtrar todas las variables irrelevantes de la base original, dejando las ventas diarias, países y canales de ventas. De esta manera, se puede separar las ventas de cada uno de los países para ajustar los algoritmos y sus parámetros a los elementos propios de cada país como feriados, eventos, entre otros.

Por otro lado, se requiere formatear los datos con la estructura de **ds: fechas** e **y: valores de ventas** para cada uno de los países, esto considerando la prueba e implementación de los algoritmos de machine learning que requieren esta estructura para los datos de entrada.

Tabla 3.4: Ejemplo de estructura de data para modelos en Sodimac

ds	y
01-01-2019	xxx
02-01-2019	xxx
...	...
30-10-2021	xxx
31-10-2021	xxx

Finalmente, propio de la implementación de modelos de machine learning, se realiza la separación de los datos de cada país en subconjuntos de entrenamiento y prueba. Para esto se define la fecha de 30 de junio del 2021 como separación, dejando la información diaria de enero-2019 a junio-2021 como conjunto de entrenamiento y la data de julio-2021 a octubre-2021 como conjunto para la prueba de desempeño de los modelos en base a las métricas explyadas en la sección 2.7.

Selección y prueba de algoritmos de machine learning

Los algoritmos a implementar para las predicciones de venta se basan en los algoritmos desarrollados por Facebook Open Soruce para pronóstico de series de tiempo: Prophet y el recientemente desarrollado Neural Prophet, ambos caracterizados por generar predicciones rápidas y con parámetros interpretables por usuarios no expertos en el machine learning.

Este tipo de algoritmos generan predicciones aprendiendo de la historia pasada y su descomposición en los componentes de una serie de tiempo: tendencia, estacionalidades, efecto de feriados y posibles otros regresores. En el caso de NeuralProphet, este es un modelo híbrido que logra unir los modelos tradicionales de descomposición con los modelos de deep learning que utilizan redes neuronales, con sus beneficios y complejidades vistas en la sección 2.6.2.2.

Bajo estas características se sustenta la selección de estos algoritmos para su prueba en las predicciones de venta diaria y mensual de Sodimac, dado que se busca utilizar algoritmos comprensibles e interpretables por los usuarios de la unidad de negocio y que no sean cajas negras sin lograr incorporar y extraer conocimiento del negocio a las predicciones mismas.

Para el ajuste de parámetros de estos modelos, se enfoca en los cuales tienen directa relevancia en los cambios de tendencia, estacionalidades y efecto de los feriados. Para Prophet, este se enfoca en:

- **changeoint_prior_scale** y **seasonality_prior_scale**: Parámetros enfocados en la tasa de cambio de la tendencia y estacionalidad respectivamente.
- **seasonality_mode**: Parámetro que asigna el modo de cambio de la estacionalidad en el tiempo, esto con respecto a la tendencia y el comportamiento de la varianza de los datos (heterocedasticidad).
- **x_seasonality**: Parámetros que permite asignar con un booleano (verdadero o falso) la presencia de estacionalidades en los datos y su efecto en el entrenamiento de los modelos.

En cuanto a NeuralProphet, este se enfoca principalmente en torno a parámetros de la red neuronal autoregresiva (AR-NET), donde el modelo permite dejar en “auto” gran parte de estos parámetros para que sean determinados por el mismo modelo en el entrenamiento. En cuanto al ajuste de la tendencia, esta se maneja mediante los puntos de ajuste o **n_changeoints**.

Los parámetros que se ajustan con mayor variación en las pruebas corresponden a:

- **seasonality_mode**: Parámetro que asigna el modo de cambio de la estacionalidad en el tiempo, esto con respecto a la tendencia y el comportamiento de la varianza de los datos (heterocedasticidad).
- **loss_func**: Función de pérdida relacionada a la red neuronal y como esta va aprendiendo en el entrenamiento del modelo de acuerdo al optimizador del descenso de gradiente estocástico.
- **num_hidden_layers**: Parámetro que controla la cantidad de capas ocultas dentro de la AR-NET, permitiendo captar dinámicas no lineales dentro de la componente autoregresiva.

A partir de las propias particularidades de cada país y su comportamiento de ventas, se definen estos parámetros en específico para cada uno de ellos mediante pruebas y validaciones cruzadas. Para ambos modelos, considerando que utiliza como base la data diaria de ventas, se fija la estacionalidad semanal y anual verdaderas.

El resumen de los parámetros para cada modelo se muestra en la tabla 3.5 para Prophet y 3.5 para el modelo Neural Prophet. En general, las grandes variaciones de parámetros se ven en el efecto de estacionalidades para Prophet y la presencia de capas ocultas en NeuralProphet.

Tabla 3.5: Parámetros para modelos de predicción de ventas en base a Prophet

Parámetros para Prophet			
País	seasonality_prior_scale	changepoints_prior_scale	seasonality_mode
Chile	0.5	0.01	additive
Colombia	0.5	0.01	additive
Brasil	0.5	0.05	additive
Uruguay	0.5	0.01	additive
México	0.5	0.01	additive
Perú	0.1	0.01	additive
Argentina	0.1	0.01	additive

Tabla 3.6: Parámetros para modelos de predicción de ventas en base a NeuralProphet

Parámetros para NeuralProphet			
País	loss_func	num_hidden_layers	seasonality_mode
Chile	huber	0	additive
Colombia	huber	0	additive
Brasil	huber	2	additive
Uruguay	huber	1	additive
México	huber	1	additive
Perú	huber	2	additive
Argentina	huber	2	additive

Por último, para añadir la naturaleza calendaria propia de cada uno de los países se añaden en cada uno de los modelos los feriados legales mediante un regresor extra de `add_country_holidays` disponible en la misma librería, la cual permite integrar la información de los feriados legales de cada país en los modelos.

3.2.2.3. Proyecto: Predicción de deuda acumulada mensual para Falabella Inmobiliaria

Definición de objetivos

El tercer proyecto de transformación digital se desarrolla en el proceso de predicción de deuda acumulada para el negocio de Falabella Inmobiliaria, y que impacta con otros procesos de la unidad de negocio como la realización de presupuestos, gestión de control y cobranza, entre otros.

A partir del información de deuda por facturas impagas, emisión de documentos, notas de crédito de rectificación, y a fin de cuentas, toda documentación relacionada al comportamiento histórico de pago y deuda de los clientes, se define como objetivo generar un modelo de predicción de deuda real acumulada para el año 2022 para Falabella Inmobiliaria, con

resolución mensual y apertura por sociedades internas existentes dentro de la unidad de negocio.

Identificación, extracción y caracterización de datos - Análisis exploratorio

Para la predicción de deuda acumulada en sus distintos niveles, se utiliza la información de partidas de documentos del negocio disponible en la plataforma de gestión de datos de SaS, la cual contiene documentación de todos los flujos de pago, notas de crédito, emisión de documentos, facturas asociadas al arriendo de locales, edificios, centros comerciales, espacios de publicidad, entre otros.

Esta información presenta una historia desde octubre de 2018 al cierre de noviembre de 2021, con 19 variables cualitativas y cuantitativas de sus movimientos contables. Entre estas variables se incluye:

- Información del clientes asociado a las transacciones, con nombre, RUT, código de cliente, sociedad a la cual pertenece el cliente y su cuenta, entre otros.
- Información relacionada a los documentos con referencias internas del negocio, tipo de documentos con su codificación (pago, nota de crédito, factura, etc.), textos de partidas, cuenta del negocio relacionada al documento, etc.
- Información de fechas relacionadas a los documentos y sus flujos como fecha de emisión, vencimiento, fechas de pago y días de demora en caso que corresponda.
- Información cuantitativa de los documentos relacionando montos de los importes, donde su signo positivo o negativo depende si son facturas, notas de crédito, saldos, etc.

Se entiende que la base la información SaS es una base dinámica con los flujos de documentos que se actualiza en tiempo real cada vez que se ingresa una factura, pago, etc., por lo cual requiere realizarse una extracción y filtración que genere una imagen estática de la base de datos de la deuda acumulada.

De esta manera, se genera una base estática con la información de deuda real consolidada a cierre de cada mes, por consecuente resolución mensual de la deuda y con apertura por negocio y las dos sociedades internas: Falabella Inmobiliaria (C069) y Open Plaza (C106).

Caracterizando los datos de deuda, un beneficio de obtener la información desde SaS es que la base se consolida y actualiza día a día con los nuevos documentos por el área de control de gestión, lo cual favorece a que no existan valores anómalos ni vacíos para la data que se posee. De igual manera se validan las curvas de deuda con conocimiento experto del área de cobranza.

Analizando el comportamiento de la deuda en el tiempo y sus patrones característicos en la figura 3.10, se logra observar como existe en general una tendencia al alza en la deuda desde fines de 2019 para el total del negocio de Falabella Inmobiliaria. Este aumento de deuda se explica en mayor medida por el estallido social y el inicio de cuarentenas por la

pandemia, ambos cuales impactaron fuertemente en el sector comercial y por consecuencia un incremento de deuda para esos meses.

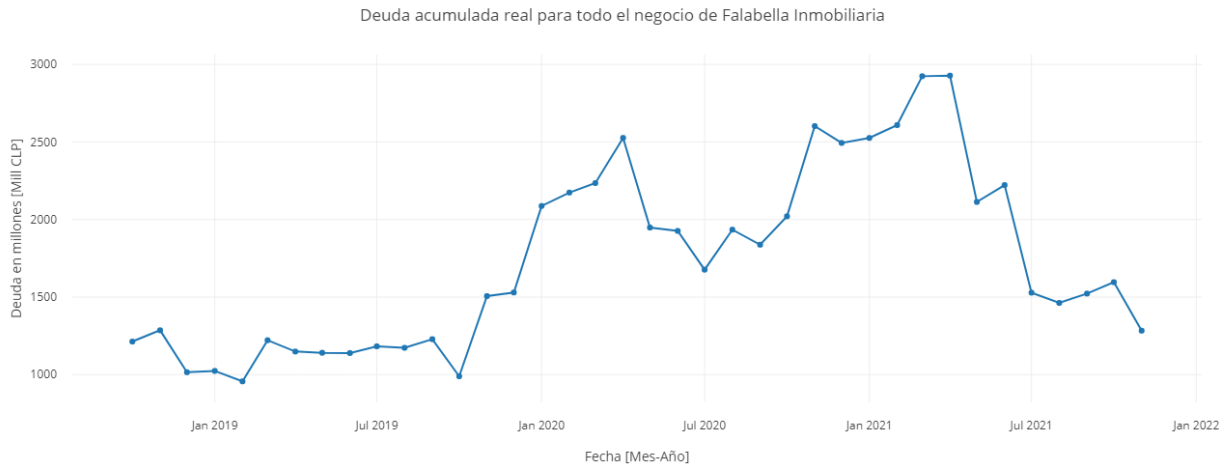


Figura 3.10: Deuda acumulada real para todo el negocio de F. Inmobiliaria

Otro patron relevante de analizar se observa para abril de 2021, donde tras llegar al máximo histórico de deuda, cambia bruscamente la tendencia con una pendiente negativa reflejada en la reducción a noviembre de 2021 de más de 1600 millones de pesos. Este comportamiento, pese a ser un cambio importante en la tendencia, tiene clara razón con los objetivos del negocio de disminuir la deuda impaga de sus clientes, realizando gestión de cobranza para esto y en general buscando no acumular deuda muy antigua de sus clientes.

Con respecto al comportamiento estacional ocurre un fenómeno similar donde ocurren incrementos de deuda acumulada en los primeros trimestres de los años y disminuciones de esta misma a mitad de año, pero que propio de la disminución brusca de deuda en los meses del 2021, se perturba su estacionalidad anual donde meses que se caracterizaban en el pasado por aumentar en deuda, para este año disminuye cuantiosamente.

Analizando las series de tiempo de deuda aperturadas por sociedad en la figura 3.11, el primer aspecto a notar es que la sociedad C106 se generó solo el pasado 2020 y por esto posee menos historia de deuda en comparación a la sociedad C069. Este aspecto indica directamente que los parámetros para los modelos a implementar en la predicción de deuda de las sociedades deberá ajustarse acorde a su data disponible.

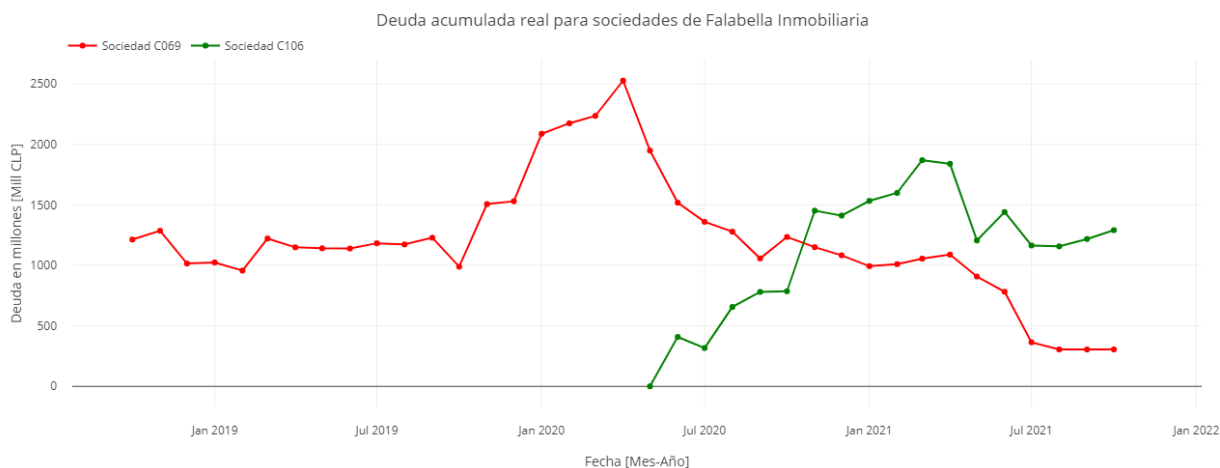


Figura 3.11: Deuda acumulada real para las sociedades de F. Inmobiliaria

En cuanto a sus patrones, para la sociedad C069 se observa una clara tendencia de disminución de deuda a partir del segundo trimestre del 2020, lo cual se opone completamente a la sociedad C106 la cual presenta un gran crecimiento en la deuda acumulada hasta julio-2021, mes donde se estabiliza la deuda de la sociedad en torno a los 1.200 millones de pesos.

Preprocesamiento de datos para algoritmos de machine learning

Para la generación de las predicciones mensuales de deuda acumulada a nivel de negocio y sociedades para 2022 se parte desde los datos estáticos a cierre de mes filtrados desde la base de SaS. A estos se les filtra todas las variables de cliente, documentos, referencias internas, y en general, toda variable con información que pudiese entregar aperturas de deuda pero que son irrelevantes por ahora.

De esta manera se posee la base en excel con 3 hojas, una para cada nivel de apertura de los datos: total del negocio, sociedad C106 Falabella Inmobiliaria y sociedad C069 de Open Plaza. A esta data, en vías de la implementación de los modelos de predicción Prophet y NeuralProphet de Facebook Open Source, se formatea de acuerdo a la estructura columnas **ds**: fecha mensual e **y**: valor de deuda mensual requerida por ambos modelos.

Tabla 3.7: Ejemplo de estructura de data para modelos en Falabella Inmobiliaria

ds	y
01-01-2019	xxx
02-01-2019	xxx
...	...
29-11-2021	xxx
30-11-2021	xxx

Finalmente, propio del trabajo con algoritmos de machine learning, se separa las curvas de deuda para las 3 aperturas en conjuntos de entrenamiento y prueba, con la data desde sus inicio hasta agosto-2021 en el conjunto de entrenamiento y para el conjunto de prueba los últimos 3 meses de la deuda: septiembre, octubre y noviembre de 2021.

Selección y pruebas de algoritmos de machine learning

En el desarrollo de las proyecciones para la deuda general del negocio y la deuda para las sociedades, se debe considerar que cada serie de tiempo posee características distintas desde más o menos historia, hasta tendencias y estacionalidades propias de cada apertura de deuda del negocio.

Debido a lo anterior, se requiere ajustar los modelos para cada nivel de apertura de manera distinta, buscando capturar correctamente su historia pasada y predecir de mejor manera la deuda a 2022 tanto para el negocio como para ambas sociedades.

A partir de lo anterior y sumado a que se busca que los algoritmos de machine learning a utilizar sean entendibles en un área sin mayor uso de tecnologías desde la inteligencia artificial, se decide seleccionar los algoritmos de machine learning de Prophet y Neural Prophet, ambos desarrollados por Facebook Open Source que se destacan por generar predicciones rápidas, flexibles y con parámetros interpretables y ajustables para usuarios no expertos, permitiendo conectar información de deuda del pasado a los mismos parámetros, que a su vez, inciden en el desempeño y ajuste de los algoritmos en la predicción de la deuda acumulada a futura.

Considerando que ambos algoritmos se basan en la descomposición de series de tiempo, ya sea en tendencia, estacionalidad, covariables, efecto de feriados, entre otros, su ajuste de parámetros se hace propio a estas componentes para cada uno de los modelo.

Los parámetros en los cuales se centró el ajuste para el modelo Prophet fue:

- **seasonality_mode**: Parámetro que ajusta la forma de cambio de la tendencia en el tiempo, con modo aditivo o multiplicativo.
- **seasonality_prior_scale** y **changepoint_prior_scale**: Parámetros de ajuste de cambio de la estacionalidad y tendencia.

En cuanto a NeuralProphet, este algoritmo no utiliza los parámetros de ajuste de cambio como **changepoint_prior_scale**, en cambio, el ajuste de tendencia se realiza principalmente mediante la cantidad de **n_changepoints** o puntos de cambio de tendencia, dejando que el mismo modelo en su entrenamiento determine la ubicación óptima de los puntos de cambio capturando de la mejor manera la historia pasada para futuras predicciones.

Para los parámetros de aprendizaje de la red neuronal, estos se fijan de manera “auto” para que el mismo modelo en su entrenamiento determine las épocas, el batch size, ratio de aprendizaje, entre otras variables.

Los parámetros que se modifican y prueban que no son automáticos son:

- **loss_func**: Función de pérdida relacionada a la red neuronal y como esta va aprendiendo en el entrenamiento del modelo de acuerdo al optimizador del descenso de gradiente estocástico.
- **n_changepoints**: Parámetro que fija la cantidad de puntos donde la tendencia cambia su comportamiento y permitiendo controlar la flexibilidad de la tendencia en cierta medida.
- **num_hidden_layers**: Parámetro que controla la cantidad de capas ocultas dentro de la AR-NET, permitiendo captar dinámicas no lineales dentro de la componente autoregresiva.
- **changepoint_range**: Asigna cuanto de la historia se utiliza de mayor manera para el ajuste de la tendencia. Este parámetro resulta relevante en las series de deuda que se marca notoriamente la disminución brusca que ocurrió el ultimo semestre del 2021, buscando ajustar correctamente la historia reciente en el aprendizaje del modelo.

Entendiendo que el ajuste es propio para cada serie de tiempo de apertura de deuda acumulada, en la tabla 3.8 y 3.9 se entrega un resumen de los parámetros finalmente seleccionados para las predicciones finales, esto tras pruebas y validaciones cruzadas sobre los datos de entrenamiento.

Tabla 3.8: Parámetros para modelos de predicción de deuda acumulada en base a Prophet

Parámetros para Prophet			
Apertura	seasonality_prior_scale	changepoints_prior_scale	seasonality_mode
Negocio General	5	0.1	multiplicative
Sociedad C069	0.1	0.01	additive
Sociedad C106	0.1	0.05	additive

Tabla 3.9: Parámetros para modelos de predicción de deuda acumulada en base a NeuralProphet

Parámetros para NeuralProphet					
Apertura	loss_func	num_hidden_layers	seasonality_mode	n_changepoints	changepoints_range
Negocio General	MAE	2	multiplicative	12	0.8
Sociedad C069	MAE	2	additive	10	0.75
Sociedad C106	huber	3	additive	15	0.8

Capítulo 4

Resultados y Análisis

En el presente capítulo se presentan los resultados obtenidos en 3 secciones, una para cada proyecto de transformación digital desarrollado en la presente memoria.

Junto a los resultados, se expresan las métricas de los algoritmos de machine learning implementados en los proyectos y se realiza un análisis de los mismos resultados, expandiendo su desarrollo a los impactos a nivel de negocio y gerencia de cada uno de los proyectos.

4.1. Predicción de ventas para Tottus Chile

En esta sección se presentan los resultados del proceso de predicción de ventas mensuales para Tottus Chile, en base al algoritmo predictivo de “Prophet” y la data de ventas de supermercados Tottus Chile desde enero 2019 a julio 2021.

En la figura 4.1 se presenta la proyección a diciembre de 2021 de las ventas netas a nivel mensual para Tottus Chile, de color rojo la predicción del modelo y azul el valor real de ventas para cada mes.

En esta predicción se logra ver como el modelo logra adaptarse a la tendencia al alza de las ventas a lo largo de los años y las estacionalidades de meses como septiembre o diciembre donde se incrementa notoriamente las ventas y para disminuir en magnitud en meses contiguos.

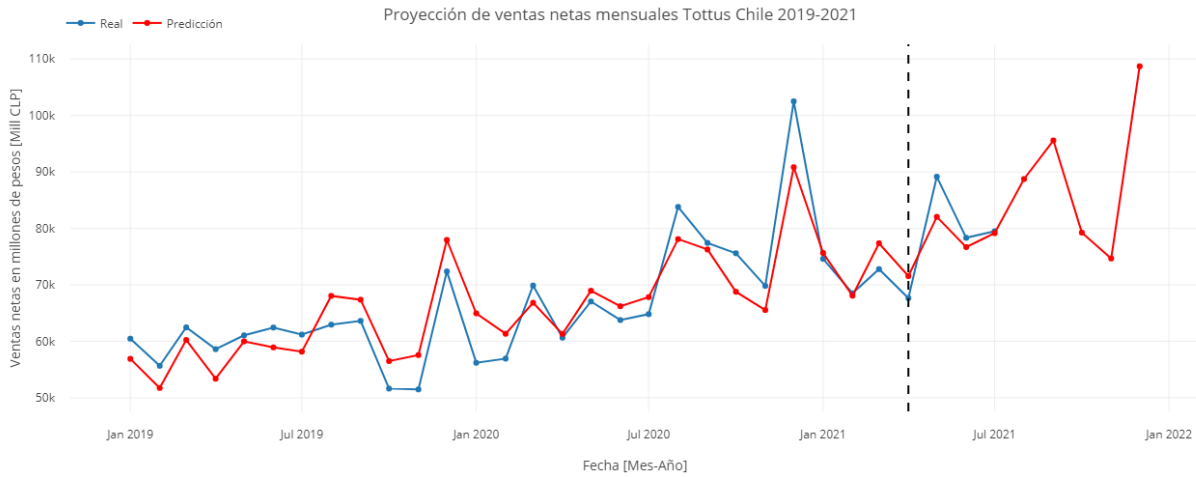


Figura 4.1: Predicción ventas mensuales para Tottus Chile

Observando en mayor detalle la predicción de ventas para el 2021 presente en la figura 4.2, se observa como la proyección de venta netas va al alza en lo que queda del 2021 hasta llegar a 108.713 millones de pesos [CLP] en el mes de diciembre.

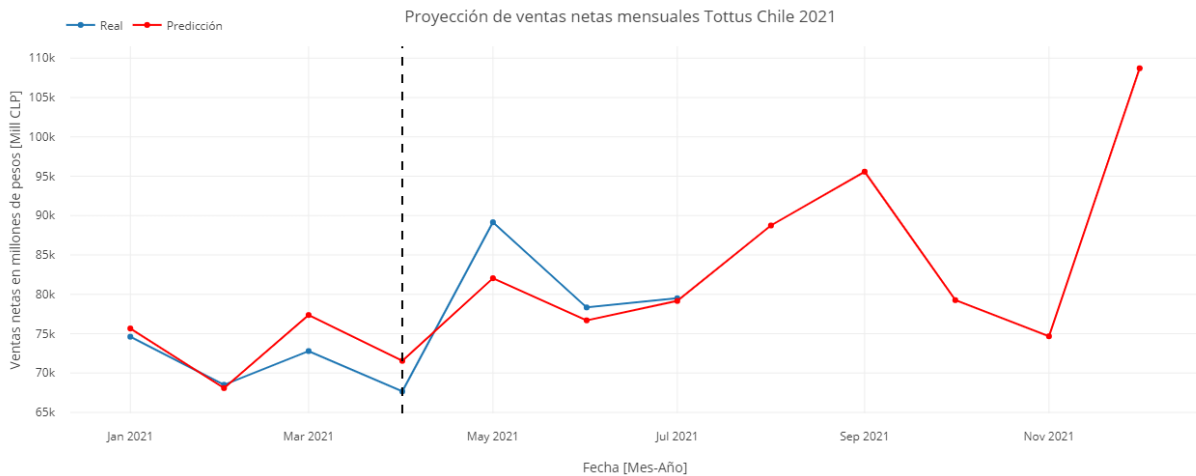


Figura 4.2: Predicción ventas mensuales para Tottus Chile 2021

Además, se logra apreciar como el modelo en el corto plazo para los meses de mayo-junio-julio posee un buen rendimiento, con un error absoluto (MAE) del 3.68 % y de 5.12 % para la raíz del error cuadrático medio (RMSE) tal cual se aprecia en la tabla 4.1. Esto representa resultados precisos, con un sesgo a sobrestimar las ventas a futuro en -3031 millones de pesos sobre el orden de 75.000 millones que tienen las ventas mensuales.

Tabla 4.1: Métricas de error en predicción mensual de ventas Tottus Chile

Métricas	Sesgo	MAE	MAE %	RMSE	RMSE %
Evaluación predicciones para may-jul 2021	-3031.04	3031.04	3.68 %	4217.65	5.12 %

Con los resultados del modelo de predicción basado en machine learning, se procede a compararlo con la metodología actual para la predicción de ventas basado en excel y reglas de negocio para el mismo periodo de los meses mayo, junio y julio del 2021.

Tal como se refleja en la tabla 4.2, el proceso actual de predicción posee un MAE a los 3 meses de 8.78 % y un RMSE de 11.71 %, ambos claramente mayores a los entregados por el modelo basado en machine learning, además de presentar un mayor sesgo de ventas, lo cual se refleja en que las predicciones mediante machine learning generan cerca de un 60 % de mejora en las métricas de desempeño.

Tabla 4.2: Comparación métricas de error predicción de ventas mensuales de acuerdo a proceso actual

Métricas	Sesgo	MAE	MAE %	RMSE	RMSE %
Forecast Excel	-7051.57	7051.57	8.78 %	9409.85	11.71 %
Forecast Prophet	-3031	3031.04	3.68 %	4217.65	5.12 %

En general, se logra apreciar que se cumple el objetivo principal de generar un modelo de predicciones de ventas mensuales para Tottus con mejor desempeño que el proceso actualmente realizado y basado netamente en datos, historia de las ventas e información propicia a las ventas minoristas.

Cabe notar, que aunque se tiene un incremento de precisión de la predicción de ventas a 3 meses, este varía cuando se incrementa el horizonte de tiempo de prueba del modelo, sin embargo, en general entrega mejores resultados en todos los casos comparativamente con el modelo actualmente utilizado.

4.1.1. Impactos y beneficios del proyecto

Junto a la mejora de precisión del modelo, la implementación de este proyecto en base a tecnologías como el machine learning y el aprovechamiento de infraestructura TI como los servicios de la nube conlleva una serie de beneficios, y fin de cuentas, impactos cuantificables de mejora en el proceso de predicción de ventas para Tottus Chile.

Automatización del proceso

Mediante el aprovechamiento de la extracción de datos transaccionales desde bigquery, el manejo de datos y la implementación del modelo de machine learning mediante Python y librerías de data science disponibles en este, se logra automatizar el proceso de predicción de ventas mensuales sin mayor interacción humana que correr un archivo python `.py`.

Esta automatización libera al proceso de horas hombre para su realización, las cuales en promedio demoraban de 5 a 10 horas mensuales pasando a menos de 5 minutos para todas las predicciones requeridas. Además, esta automatización permite desprenderse del trabajo en excel, la validación manual de datos, introducción de reglas de negocios de Tottus y por ende, más demoras del proceso.

Confiabilidad del proceso

Una derivada propia de la automatización del proceso consiste en la confiabilidad que adquiere el proceso de proyecciones de ventas para Tottus Chile. Gracias a condensar gran parte del trabajo en Python mediante reglas de preprocesamiento claras y objetivas, se evita la intervención humana en un proceso reiterativo en la unidad de negocios, y con esto, la posibilidad de introducir errores no voluntarios por la acción humana de manejar gran cantidad de datos en planillas de excel.

De esta manera, se orienta el proceso desde uno basado en reglas de negocio y conceptos subjetivos desde otras áreas de Tottus a un enfoque netamente basado en datos e historia pasada, con reglas claras, automatizado y evitando errores por manejo de datos reiterativo.

Extensión del modelo a nivel diario

Una tercera arista proviene desde la resolución original de los datos de ventas de Tottus, los cuales vienen en resolución diaria. Actualmente en la unidad de negocio solo se considera la predicción de ventas a nivel mensual, pero gracias a la versatilidad del modelo implementado, este se puede extender para generar predicciones a nivel diario.

En la figura 4.3 se aprecia de color rojo la proyección de ventas netas diarias para Tottus Chile desde mayo 2021 hasta 31 de diciembre del mismo año. En esta se logra observar correctamente el ajuste de la proyección a las estacionalidades de los meses y la tendencia a lo largo del tiempo junto a su comparación de desempeño en toda la historia disponible.

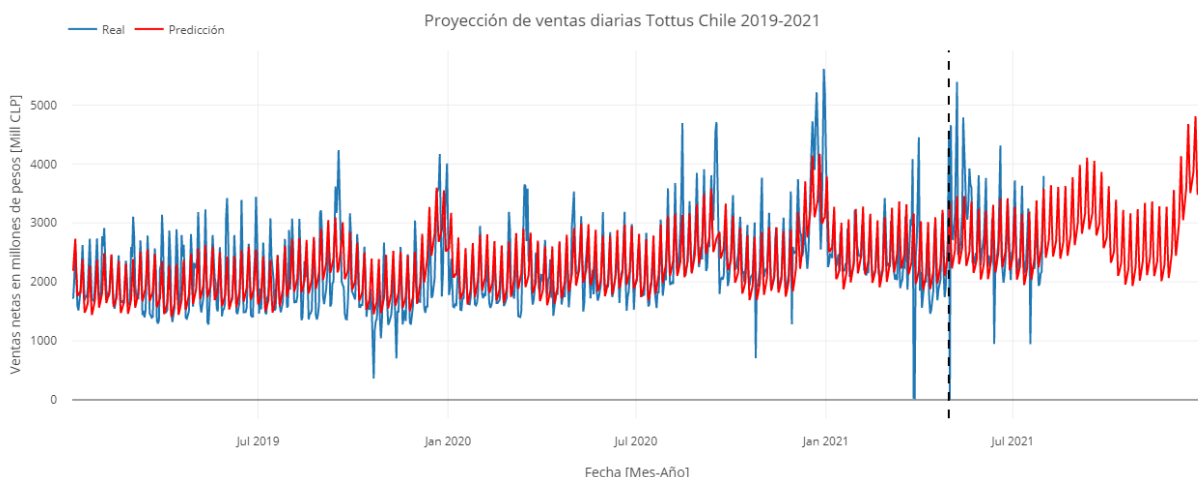


Figura 4.3: Predicción ventas diarias Tottus Chile

Para el análisis de ventas del presente año ilustrado en la figura 4.4, se nota que las ventas a nivel diario poseen una clara estacionalidad anual con proyección de incremento de ventas para septiembre y diciembre, fechas claramente caracterizadas por ser buenas en ventas por fiestas patrias y fiestas de fin de año.

En el detalle de la resolución diaria, se observa que pese a añadir los efectos de feriados legales y fechas connotadas del último año como retiros de AFP, existen peaks y disminuciones no consideradas que se separan del comportamiento general de las ventas netas para Tottus Chile. Esta diferencia claramente incide en la precisión del modelo, ya que se acumula error al no lograr abarcar esos peaks de ventas.

Analizando estos días, se advierte que coinciden con fechas comerciales características del retail en el país como cybermondays, blackfridays, día de la madre, semanas de ofertazos, entre otras, lo cual señala una clara arista de mejoras al modelo de predicción en resolución diaria.

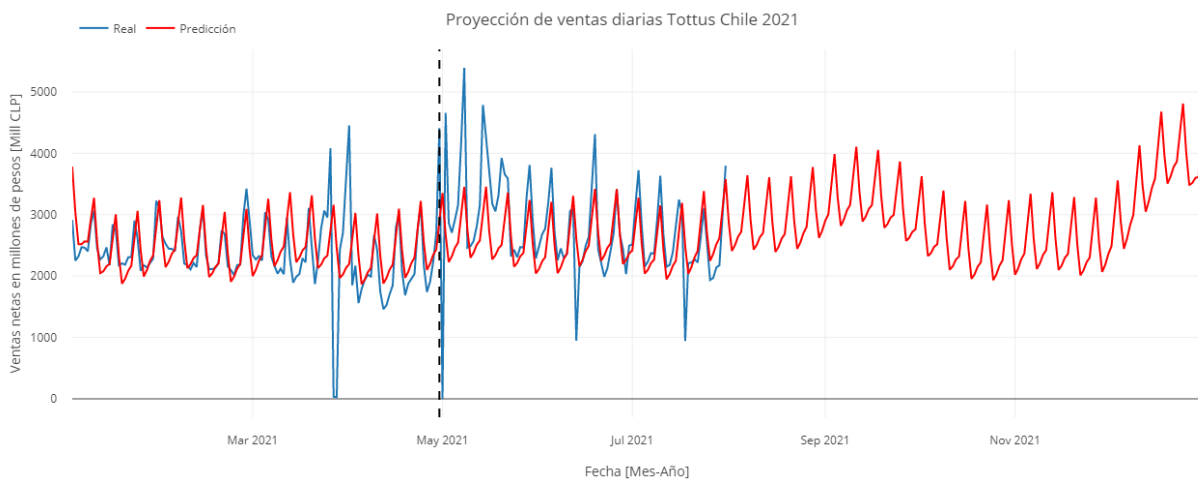


Figura 4.4: Predicción ventas diarias Tottus Chile 2021

Este efecto a nivel diario se valida con las métricas de error del modelo Prophet sobre las predicciones, como se observa en la tabla 4.3, donde el error absoluto (MAR) y la raíz del error cuadrático (RMSE) se incrementa a un 26.2% y 33.6% respectivamente, ambas claramente mayores a las métricas de error del modelo en resolución mensual.

Tabla 4.3: Métricas de error en predicción diaria de ventas Tottus Chile

Métricas	Sesgo	MAE	MAE %	RMSE	RMSE %
Evaluación sobre el test-set	-175.45	718.99	26.20 %	922.71	33.62 %

Extensión del modelo mensual a supermercados

Al igual que la extensión del modelo a nivel diario, dada la capacidad e información disponible en la base transaccional disponible en bigquery de GCP, se permite generar predicciones mensuales para las sucursales en funcionamiento de Tottus Chile.

Considerando que actualmente Tottus posee 69 locales operativos, ilustrar la predicción para cada uno de estos se volvería un ejercicio engorroso, por lo cual, se procede a ejemplificar la extensión del modelo a predicción de ventas mensuales en sucursales con el local de Los Dominicos para este 2021.

En la figura 4.5, se logra apreciar la predicción de ventas con un comportamiento similar a las proyecciones mensuales de todo Tottus, donde se ajusta correctamente a la tendencia incremental de ventas a lo largo de los meses y variaciones propias de la estacionalidad mensual la cual da el aspecto de sierra.

Este último aspecto se traduce en las proyecciones de ventas para septiembre de 1466 [Mill CLP] y diciembre de 1802 [Mill CLP], meses caracterizados por ser de los mejores meses para las ventas minoristas y que sumados a la tendencia al alza de ventas se proyectan de tal manera.

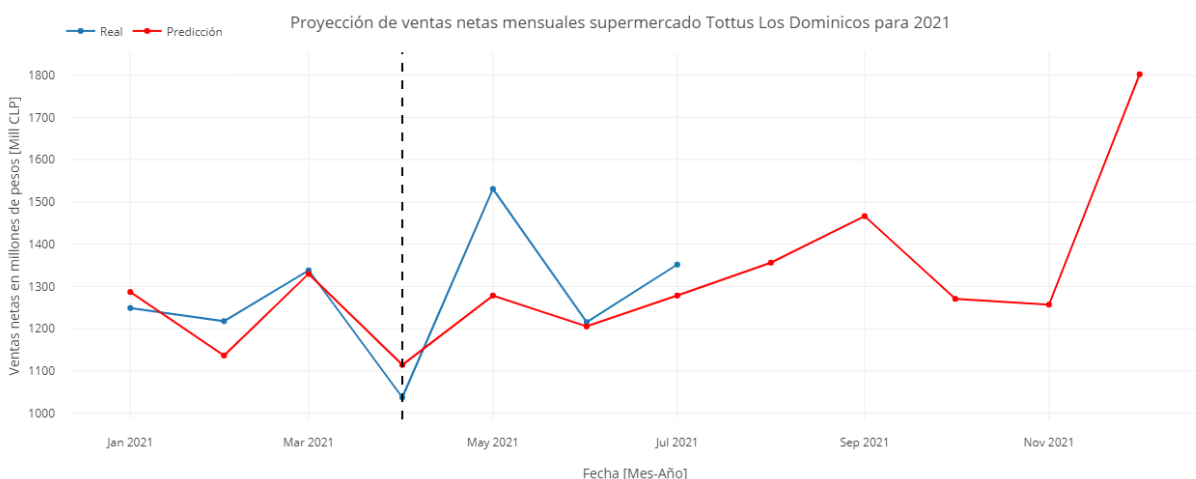


Figura 4.5: Predicción ventas mensuales para local de Los Dominicos Tottus 2021

Analizando el desempeño del modelo implementado, se procede a comparar los resultados con la desagregación que se realiza en el proceso de proyección basado en excel, el cual mediante reglas de negocio descompone la venta general de Tottus Chile en las ventas de sus supermercados.

Como se observa en la tabla 4.4, el nuevo modelo para predicción de ventas nuevamente genera mejores resultados que su contra parte, con una reducción del MAE de 9.81 % a 8.2 %. En cambio para el RMSE se presenta un comportamiento inverso donde el error del modelo en base a excel presenta una métrica ligeramente menor.

Tabla 4.4: Comparación métricas de error predicción de ventas mensuales para local de Los Dominicos

Métricas	Sesgo	MAE	MAE %	RMSE	RMSE %
Forecast Excel	-34	133	9.81 %	138	10.16 %
Forecast Prophet	-112	112	8.20 %	151.86	11.12 %

Este comportamiento se explica donde para el mes de mayo, el modelo de machine learning se separa en cerca de 250 millones con respecto al valor real y la métrica de RMSE castiga con mayor peso esta diferencia de error para el nivel de ventas del mismo mes. A su vez, esta

diferencia se explica por fechas comerciales del mes como el día de la madre junto al comienzo de entrega del tercer retiro de AFP, lo que incrementó el dinero circulante potencialmente utilizado para compras. Este análisis da a entender que el modelo implementado aun tiene campo de mejora en su desarrollo futuro.

De igual manera, considerando los beneficios de automatización, donde ejecutar el modelo mensual para todos los locales no toma más de 5 minutos, la precisión y confiabilidad del modelo en base a machine learning y otras características, destacan estos resultados para utilizarlos sobre el modelo en base excel.

Expansión modelo a otras unidades de negocio

Como última arista de impacto del modelo de predicción de ventas para Tottus se desliza en la expansión y replicación del modelo a otras unidades de negocios con características de venta minorista como Sodimac Latam.

A partir del desarrollo en el modelo de machine learning implementado en Tottus Chile, sus resultados y la inexistencia actual de un modelo de predicción de ventas en Sodimac, se procede a replicar el trabajo en esta unidad de negocio.

Esto se alinea en que el trabajo analítico en base a datos y tecnologías digitales que muestre resultados cuantificables e impacto en procesos genera señales e interés para otras unidades de negocio de involucrarse y sumarse en los proyectos focalizados en transformación digital.

4.2. Predicción de ventas para Sodimac Latam

En la presente sección se desarrollan los resultados para la predicción de ventas a nivel diario y mensual de Sodimac en los 7 países que posee presencia: Chile, Colombia, Argentina, Brasil, Uruguay, México y Perú.

Este análisis se presenta en primer lugar desagregado por países con la prueba de ambos modelos desarrollados: tanto Prophet como NeuralProphet, comparando sus métricas de desempeño considerando que no existe un proceso actual en la unidad de negocio que permita contrastar los resultados de las predicciones de venta.

Posteriormente, se exponen brevemente los resultados de proyecciones de ventas mensuales y diarias del mejor modelo en cada uno de los países, agrupados para todo Sodimac Latam para el último trimestre de 2021.

4.2.1. Pronóstico de venta diaria

En la presente subsección se presentan los resultados de predicción de ventas diarias para cada uno de los países que posee presencia Sodimac en latinoamérica y las métricas de desempeño de los modelos de machine learning implementados.

Chile

En la figura 4.6 se presentan los resultados de la predicción de ventas diarias hasta el cierre de año 2021, con la vertical negra que separa los últimos 3 meses donde se validan los resultados a partir de las métricas de error.

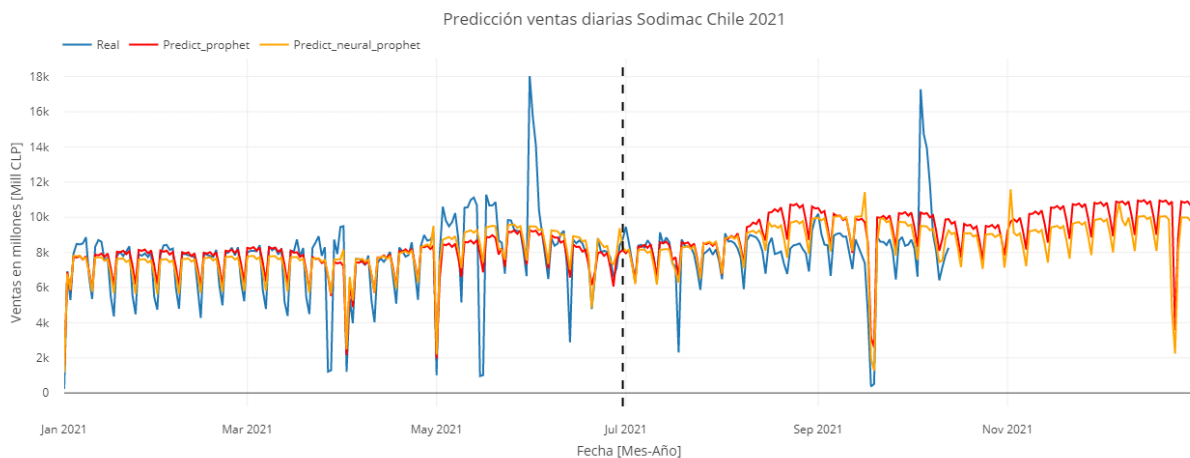


Figura 4.6: Predicción ventas diarias - Sodimac Chile 2021

Se logra apreciar como las predicciones de los modelos capturan la tendencia al alza en largo plazo lo cual permite proyectar ventas del orden de 10.000 millones de pesos para diciembre del 2021.

Comparativamente, el modelo en base a NeuralProphet proyecta las ventas en menor grado que Prophet y logra capturar de mejor manera las estabilización de ventas propio de la estacionalidad semanal, donde los viernes tienden a disminuir comparado a otros días de la semana y genera las puntas hacia abajo.

Sin embargo, ambos modelos no logran ajustarse a peaks puntuales como el ocurrido el 4, 5 y 6 de octubre donde se celebró el cybermonday en Chile y las ventas llegaron al nivel de 17.000 millones de pesos, un incremento de ventas claramente no predecible por no ocurrir en el pasado.

Pese a estos eventos puntuales de fechas comerciales, ambos modelos logran buenos desempeños los cuales se reflejan en las métricas de error de la tabla 4.5, con sesgos menores a los 1000 millones y errores menores al 16 %.

Tabla 4.5: Métricas predicción ventas diarias Chile

Métricas Mensuales Chile	Sesgo	MAE	MAE %	RMSE	RMSE %
Prophet	877.10	1281.11	15.72 %	1685.27	20.68 %
Neural Prophet	477.07	955.67	12.22 %	1506.44	18.49 %

Comparativamente, al igual que lo ilustrado en la figura de venta diaria, se observa como el modelo en base a NeuralProphet entrega ligeramente un mejor desempeño con un error absoluto MAE % de 12.22 % versus el 15.72 % del modelo Prophet. Este desempeño, sumado a que posee un sesgo 400 millones menor y un RMSE también menor, refleja que el modelo NeuralProphet obtiene la mejor predicción de ventas a fin de año para Sodimac Chile.

Colombia

Se procede con los resultados de las predicciones diarias para Sodimac Colombia. En la figura 4.7 se muestra la venta real más las predicciones de ventas de los modelos Prophet y NeuralProphet, donde ambas logran captar la tendencia y sus cambios a lo largo de la historia de ventas de Sodimac.

Además se aprecia la clara estacionalidad anual de las ventas, donde para el mes de diciembre aumentan las ventas en comparación a enero y noviembre, lo cual se refleja en la forma de cerro de la predicción para fin de año.

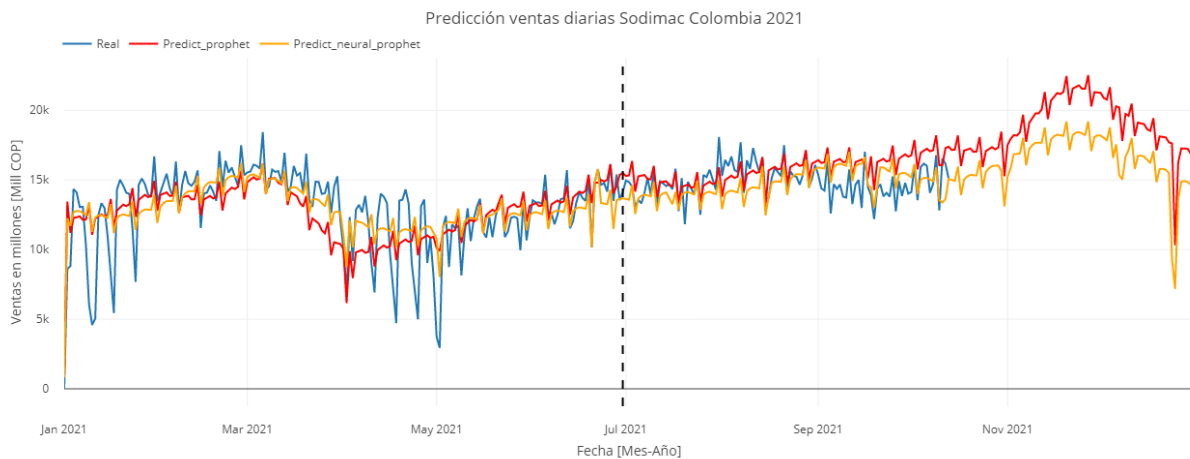


Figura 4.7: Predicción ventas diarias - Sodimac Colombia

Analizando los resultados Prophet como NeuralProphet para el último trimestre de 2021, se logra apreciar como ambos modelos reflejan correctamente el efecto de la estacionalidad anual de las ventas, sin embargo, el modelo NeuralProphet presenta una menor predicción propia de la reducción de ventas que se refleja en el mes de septiembre en la curva azul.

Esta menor predicción para diciembre y mayor cercanía a la curva real para el mes de septiembre se refleja en las métricas de error de la tabla 4.6, donde nuevamente el modelo en base a NeuralProphet presenta mejor MAE y RMSE en cerca de 1 %. La mayor diferencia se ve en el sesgo o error promedio de los modelos, donde NeuralProphet entrega un sesgo solamente de 50 millones [COP], lo cual es un monto pequeño considerando que las ventas diarias son del orden de 10.000-15.000 millones de [COP].

Tabla 4.6: Métricas predicción ventas diarias Colombia

Métricas Diarias Colombia	Sesgo	MAE	MAE %	RMSE	RMSE %
Prophet	997.03	1287.37	8.65 %	1581.62	10.63 %
Neural Prophet	50.30	1128.75	7.59 %	1430.74	9.62 %

A partir de estas métricas, se selecciona a NeuralProphet como el modelo desarrollado que genera las predicciones de venta diaria más precisas para Sodimac Colombia para el compilado final de resultados.

Argentina

En la figura 4.8 se aprecian los resultados de las predicciones de ventas diarias para Sodimac Argentina. A grandes rasgos se observa como las predicciones de venta se ajustan a la tendencia y con estacionalidad anual clara, donde meses como abril suelen tener bajas de ventas que se recuperan en meses como enero o noviembre llegando al orden de los 30 millones [ARS].

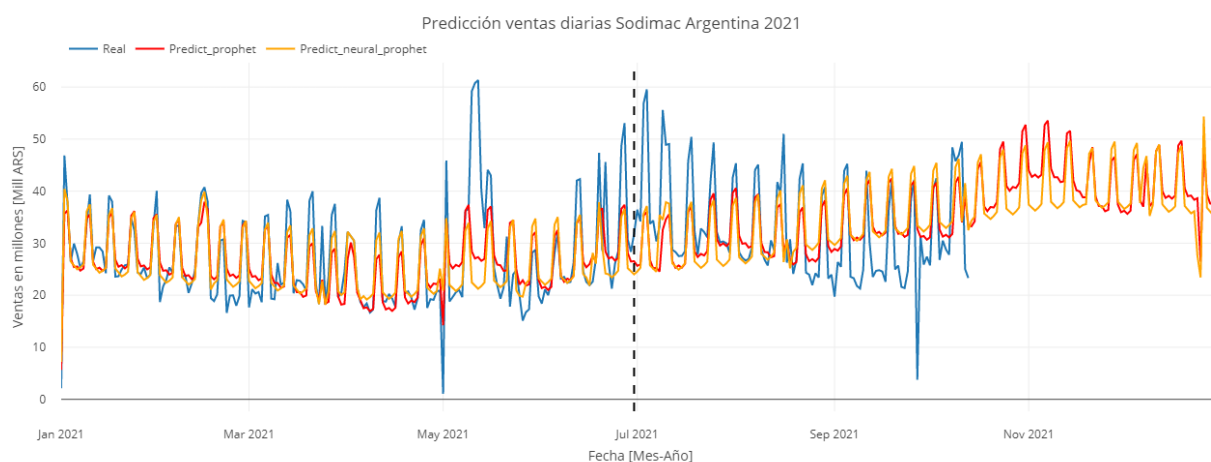


Figura 4.8: Predicción ventas diarias - Sodimac Argentina

Con mayor foco en las predicciones a fin de 2021, se aprecia de mejor manera la estacionalidad semanal de las ventas con forma de pico, aunque ninguno de los dos modelos logra captar completamente la magnitud de estos. De igual manera, ambos modelos proyectan incremento de ventas al orden de los 30 - 40 millones de pesos argentinos.

En este sentido, a partir de las métricas de error disponibles en la tabla 4.7, se observa como el modelo Prophet logra tener mejor desempeño pese a que en general las métricas para ambos están del orden de 15 % a 20 %.

Tabla 4.7: Métricas predicción de ventas diarias - Argentina

Métricas Diarias Argentina	Sesgo	MAE	MAE %	RMSE	RMSE %
Prophet	-0.49	6.04	18.48 %	8.16	24.96 %
Neural Prophet	1.00	6.97	21.33 %	8.92	27.28 %

Prophet presenta un mejor desempeño principalmente por un MAE de 6.04 [mill ARS] versus el MAE de NeuralProphet que llega a 6,97 [mill ARS]. Esto se explica a que el modelo Prophet logró capturar de mejor efecto la estacionalidad anual, donde ocurre una disminución de ventas en septiembre del 2021. Además, presenta un mejor sesgo de error promedio y un menor RMSE en cerca de 3 %, lo cual indica que sus errores son menores en proporción a las venta para ese día.

Se determina utilizar la predicción en base al modelo Prophet, lo cual muestra a que no necesariamente el uso de un algoritmos más complejos traerá mejores resultados en las predicciones de ventas. Esto se suma a que existe una dimensión de los datos con campo de mejora como fechas comerciales, eventos y contexto del mismo país y en general cualquier otro elemento cuantificable que se relacione con la venta minorista y pudiese ser integrado a los modelos de predicción de venta como covariables.

Brasil

Se procede con la predicción de ventas diarias para Brasil, la cual se refleja en la figura 4.9 donde se logra apreciar la adaptación y aprendizaje del modelo desde la historia de ventas de 2019 en Sodimac Brasil.

Se observa como las ventas en Brasil se caracterizan por fuertes peaks de baja en ventas propio de la estacionalidad semanal, donde los domingos tienden a disminuir fuertemente para recuperarse los días entre semana consecutivos.

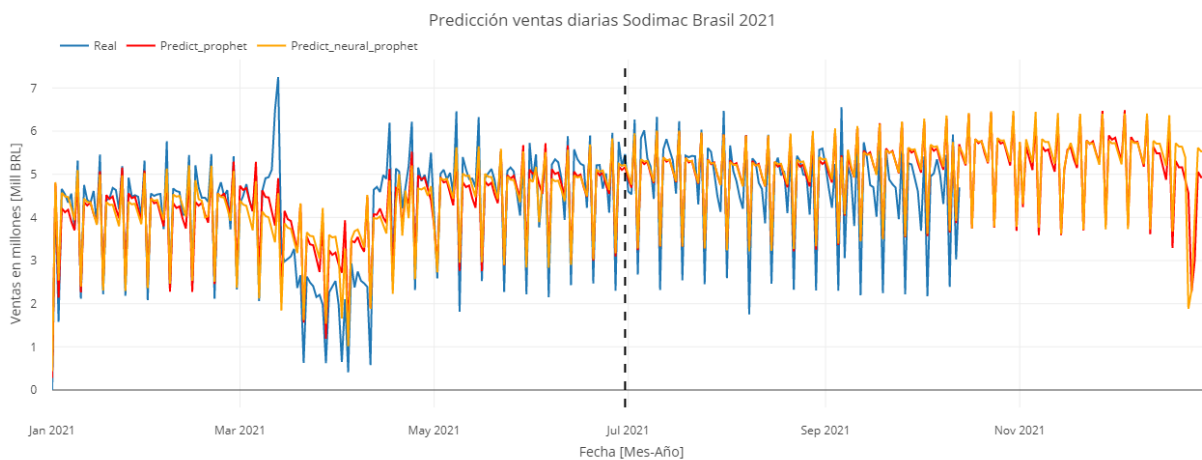


Figura 4.9: Predicción ventas diarias - Sodimac Brasil 2021

Además, se logra apreciar como las ventas capturan la tendencia al alza, superando los 5 millones de reales brasileños [BRL] para noviembre del 2021. De la misma manera, se observa como Prophet captura ligeramente de mejor manera la estacionalidad semanal de las ventas en comparación al modelo basado en Prophet, con las bajadas rojas de las proyecciones.

Este análisis se valida con las métricas de desempeño de la tabla 4.8, donde se observa como Prophet obtiene sutilmente mejores métricas, con diferencias de 0.3 % para el MAE % y 0.34 % para el RMSE %. En general, ambos modelos presentan buen desempeño para las predicciones de ventas de Brasil, pero se decide escoger para las predicciones finales a Prophet por sus mejores métricas de error.

Tabla 4.8: Métricas predicción ventas diarias Brasil

Métricas Diarias Brasil	Sesgo	MAE	MAE %	RMSE	RMSE %
Prophet	0.34	0.51	10.83 %	0.63	13.35 %
Neural Prophet	0.37	0.52	11.13 %	0.64	13.69 %

Uruguay

En la figura 4.10 se muestran los resultados de las predicciones de venta tanto para el modelo de Prophet como para NeuralProphet, en especial reflejando como se adapta a la tendencia en alza que muestra los últimos meses.

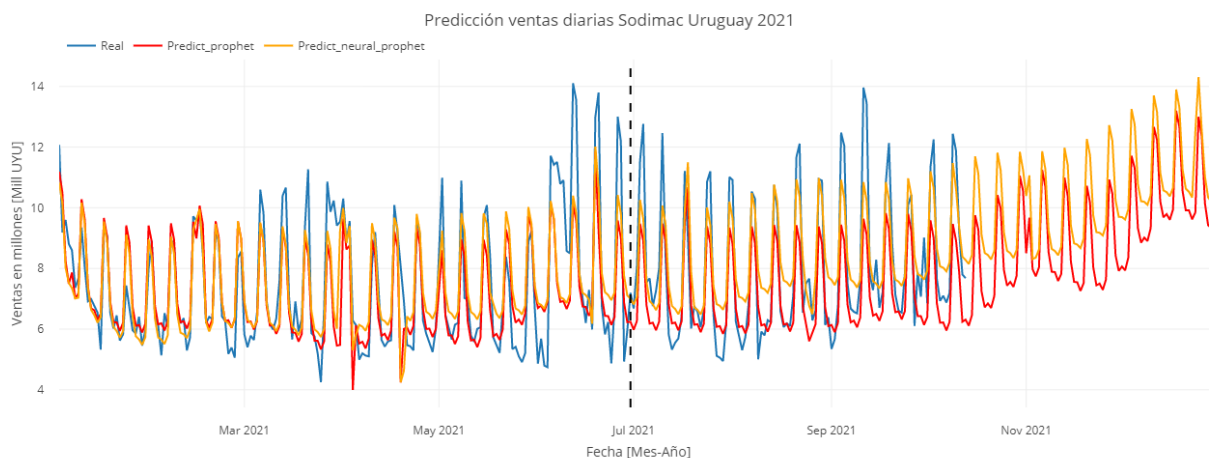


Figura 4.10: Predicción ventas diarias - Sodimac Uruguay

Se observa como ambos modelos captan correctamente los peaks propios de la estacionalidad semanal de las ventas, donde para los días entre semana (martes a jueves) tiende a disminuir las ventas y recuperarse para el fin de semana.

De esta manera proyecta las ventas diarias para el último trimestre del 2021, ambos modelos predicen ventas sobre los 10 millones de pesos uruguayos [UYU] y con peaks para los días de semana que superan los 12 millones [UYU].

Comparando el desempeño entre ambos modelos en base a sus métricas de la tabla 4.9, se muestra como el modelo en base a NeuralProphet con capas ocultas activas en la red neuronal autoregresiva logra mejores métricas de error, con un sesgo de 0.24 millones y un MAE de 12.76 %, 1.71 % menor al MAE del modelo Prophet.

Tabla 4.9: Métricas predicción ventas diarias Uruguay

Métricas Diarias Uruguay	Sesgo	MAE	MAE %	RMSE	RMSE %
Prophet	-0.86	1.16	14.47 %	1.52	18.90 %
Neural Prophet	0.24	1.02	12.76 %	1.26	15.73 %

En todas las métricas de error el modelo basado en NeuralProphet obtiene mejores métricas y visualmente se aprecia como logra captar de mejor manera los peaks semanales, por lo cual se escoge como la configuración para las predicciones de venta de Sodimac Uruguay.

México

En la figura 4.11 se muestran los resultados obtenidos para la predicción de ventas de Sodimac México. En esta se logra apreciar la leve pero consistente tendencia al alza en las ventas a lo largo de los años, con disminución de la magnitud post pandemia de covid por los claros efectos económicos que ha conllevado.

Además, se observa como ambos modelos captan la estacionalidad semanal propia de las ventas en México, donde su forma de “cerrito o sombrero” se explica por el incremento de ventas para fin de semana.

Sin embargo, para el mes de septiembre en adelante ambos modelos no capturan un incremento visible de ventas, el cual se explica por la apertura de la octava tienda en el país, el cual es un evento impredecible desde la historia pasada.

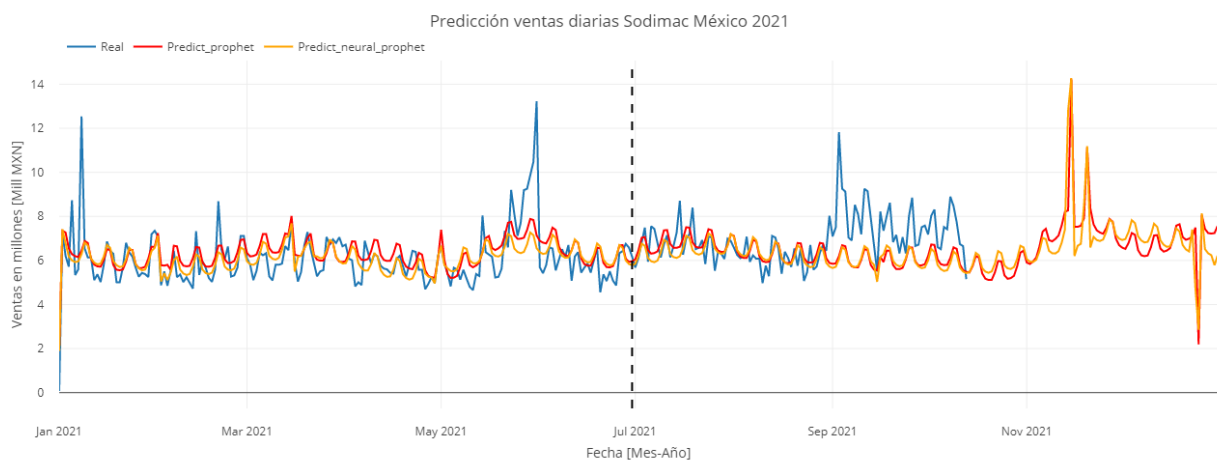


Figura 4.11: Predicción ventas diarias - Sodimac México

De esta manera, se proyectan ventas en torno a los 7 millones de pesos mexicanos [MXN] para fin de año, con una pequeña alza en las ventas para el mes de diciembre pero que no dista en mucha magnitud a los meses anteriores.

En cuanto a la diferencia de desempeño entre ambos modelos, se puede apreciar en las métricas de la tabla 4.10 como nuevamente Prophet obtiene menores métricas de error de predicción en su comparación a los meses de julio a septiembre del 2021.

Tabla 4.10: Métricas predicción ventas diarias México

Métricas Diarias México	Sesgo	MAE	MAE %	RMSE	RMSE %
Prophet	-0.6	0.97	14 %	1.30	18.74 %
Neural Prophet	-0.73	0.99	14.25 %	1.35	19.39 %

En estas, se observa como el MAE % es un 0.25 % menor para el modelo Prophet y un 0.65 % para el RMSE, ambos muy ligeramente menores. La métrica donde se presenta mayor diferencia es en el error promedio o sesgo, el cual para Prophet es de -0.6 millones versus los -0.73 millones del modelo NeuralProphet. Esto converge en escoger los resultados del modelo Prophet para la comparación final de predicciones para todos los países de Sodimac.

Perú

Se procede con la predicción de ventas diarias para el último país con presencia de Sodimac, Perú, las cuales se reflejan en la figura 4.12. En esta se aprecia a primera vista como el modelo de NeuralProphet logra captar de mejor manera la estacionalidad semanal de las ventas, y en general, produce mejores predicciones para fines de 2021.

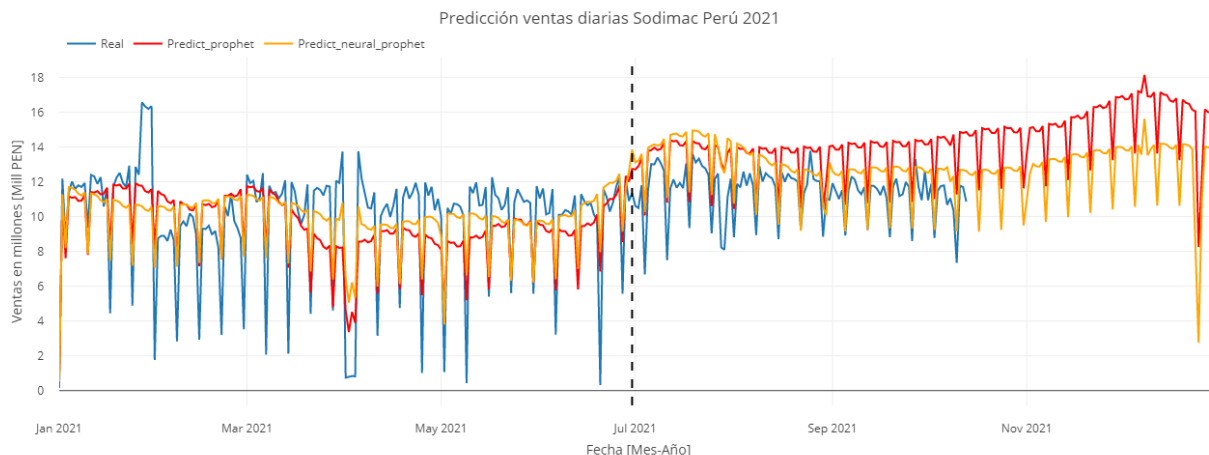


Figura 4.12: Predicción ventas diarias - Sodimac Perú

Esta estacionalidad semanal, que tiene gran importancia en la resolución diaria de las ventas, se caracteriza por una disminución considerable para los días domingos reflejándose en la punta negativa que se repite todas las semanas. Con respecto a la tendencia de ventas,

tras la atenuación visible del primer semestre de 2021, ambos modelos se ajustan a este efecto en especial para las predicciones a septiembre.

De esta manera, en la predicción para fin de 2021 se proyectan ventas en torno 13 millones de soles [PEN], donde NeuralProphet presenta menor predicción a la del modelo Prophet, pero que a su vez hace más sentido considerando las venta de los últimos meses en Sodimac Perú.

Este mejor desempeño visual del modelo NeuralProphet se sustenta en las métricas de error de la tabla 4.11, donde se aprecia que este modelo logra en todas las métricas un mejor desempeño. En el error absoluto (MAE), se obtiene un 10.53 % versus el 18.88 % del modelo Prophet, una diferencia de 8.35 % lo cual marca claramente el mejor desempeño del modelo en base a la AR-NET.

Tabla 4.11: Métricas predicción ventas diarias Perú

Métricas Diarias Perú	Sesgo	MAE	MAE %	RMSE	RMSE %
Prophet	2.15	2.15	18.88 %	2.32	20.37 %
Neural Prophet	1.11	1.20	10.53 %	1.51	13.25 %

De igual manera, se debe considerar que ambos modelos tienden a sobre-estimar las proyecciones de venta lo cual se refleja en el sesgo positivo de ambos modelos. Se desprende que, pese al desempeño mejor de NeuralProphet, aún existe campo de mejora de las predicciones desde incluir covariables con información de fechas comerciales, o que logren explicar la tendencia atenuada de ventas para los últimos meses.

Compilado de predicción de ventas diarias Sodimac Latam.

A partir de los resultados del mejor modelo para las predicciones diarias de cada uno de los países, se procede a reunir los resultados de las proyecciones de manera de tener una idea visual de las magnitudes de venta para fines de 2021.

En la figura 4.13, se logra apreciar las proyecciones con el mejor modelo para Chile y Colombia, donde ambos países siguen con grandes ventas de su moneda local llegando a diciembre con ventas en torno a los 10.000 [Mill CLP] y 15.000 [Mill COL] respectivamente.

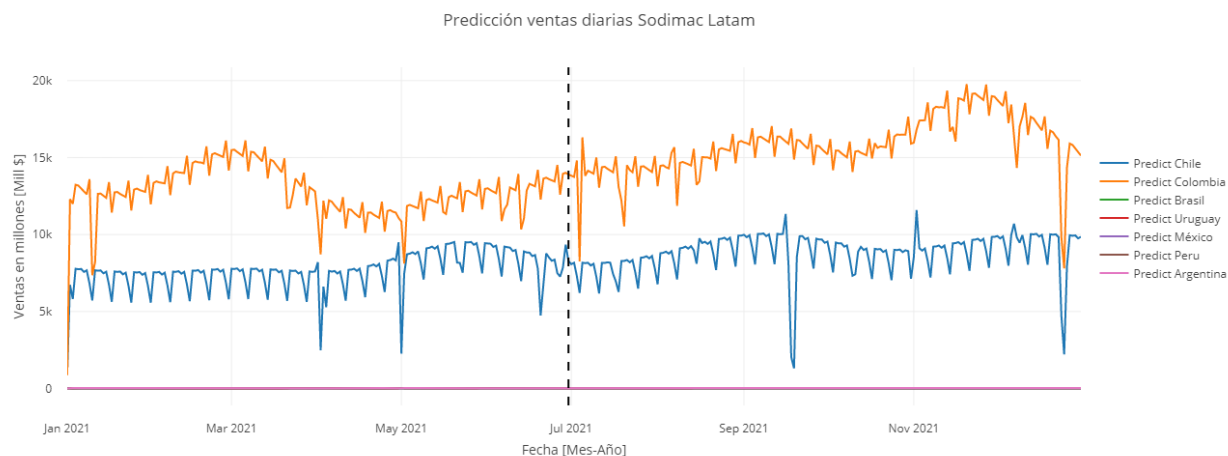


Figura 4.13: Predicción ventas diarias Sodimac Latam 1

Filtrando las ventas de Chile y Colombia que disminuyen la resolución del resto de países, se observa en la figura 4.14 de mejor manera las ventas de Argentina, características por un gran peak para el domingo y lunes de las semanas, y que en general se proyectan para fin de año en torno a los 40 millones de pesos argentinos [ARS].

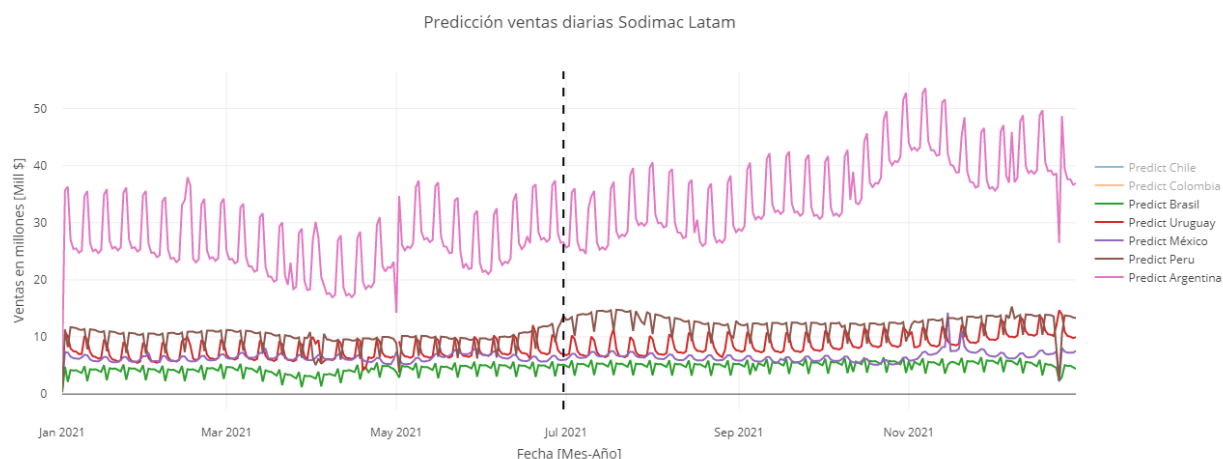


Figura 4.14: Predicción ventas diarias Sodimac Latam 2

Finalmente se refleja la proyección de ventas para los países con menor nivel de sus monedas locales como Brasil, Uruguay, México y Perú en la figura 4.15, todas claramente diferentes debido a las estacionalidades propias de cada país. De esta manera se predicen ventas en torno a los 4 [Mill BRL] para Brasil, 10 [Mill UYU] en Uruguay, 7 [Mill MXN] para México y 13 [Mill PEN] de Perú.

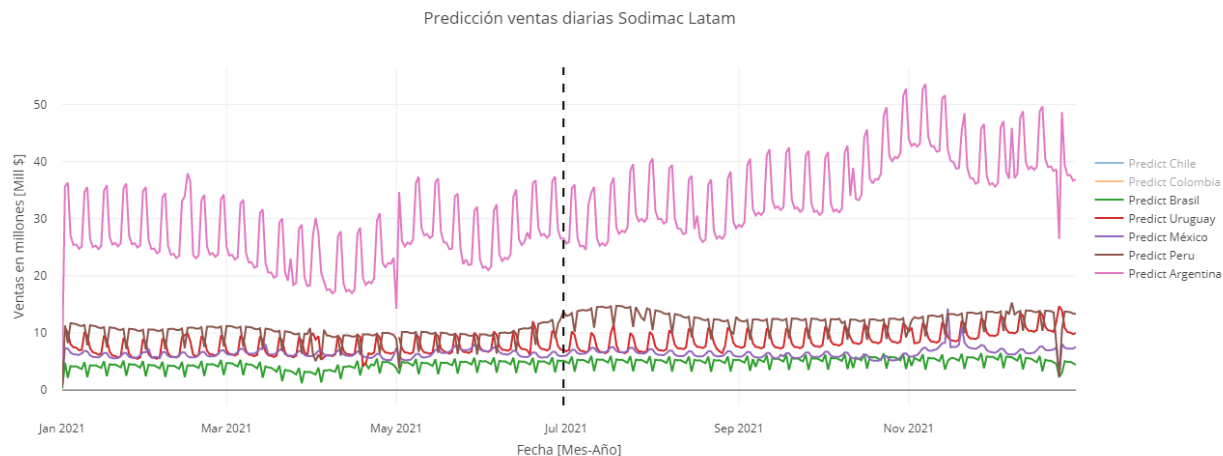


Figura 4.15: Predicción ventas diarias Sodimac Latam 3

4.2.2. Pronóstico de venta mensual

En la presente subsección se presentan los resultados de predicción de ventas mensuales Sodimac, separado por los 7 países donde posee presencia en latinoamérica.

Chile

Para la predicción de ventas mensuales, la cual tiene gran relevancia en procesos como presupuesto y control de ventas, se observa en la figura 4.16 el pronóstico de ventas mensuales para fin de 2021.

Se logra apreciar como el nivel de ventas se caracteriza por la tendencia al alza y capturando la estacionalidad anual propia de meses como mayo, septiembre y diciembre que suelen incrementar sus ventas con respecto a sus meses vecinos.

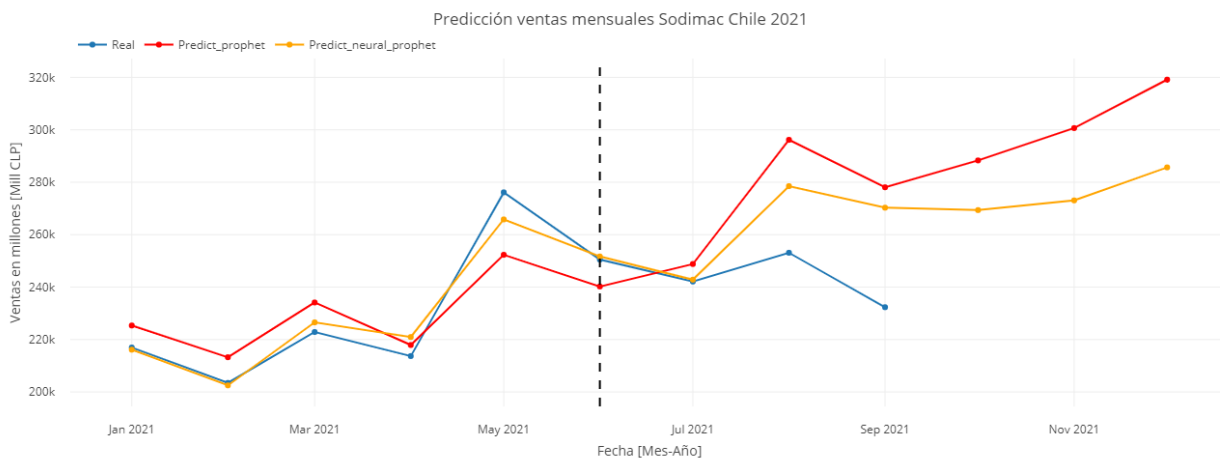


Figura 4.16: Predicción ventas mensuales - Sodimac Chile

Se observa como ambos modelos capturan la estacionalidad y tendencia de las ventas, sin embargo, el modelo en base a NeuralProphet tanto en el periodo de entrenamiento como en los meses de prueba post junio 2021 se acerca más a las ventas reales de la curva azul. De esta manera el modelo NeuralProphet proyecta ventas que llegan a los 270.000-280.000 millones de pesos para el último trimestre del 2021, menores a las de Prophet, pero que se ajustan de mejor manera a la disminución de ventas que se observa principalmente en septiembre del presente año.

Analizando las métricas de desempeño de la tabla 4.12, se valida el hecho que NeuralProphet obtiene menor error en todas las métricas, con un 8.82% de error absoluto (MAE) en comparación a Prophet que obtiene un valor sobre el 10%.

Tabla 4.12: Métricas predicción ventas mensuales Chile

Métricas Mensual Chile	Sesgo	MAE	MAE %	RMSE	RMSE %
Prophet	31835.97	31835.97	13.13 %	36482.42	15.04 %
Neural Prophet	21383.63	21383.63	8.82 %	26307.83	10.85 %

Cabe notar que ambos modelos generan un error promedio (sesgo) positivo, indicando que ambos sobrestiman las ventas a futuro, pero que en el caso de NeuralProphet no supera el 10% con respecto a las ventas mensuales reales, lo cual refuerza su selección para las predicciones de venta mensual final de Sodimac Chile.

Colombia

En la figura 4.17 se muestran los resultados para la predicción de ventas mensuales en Sodimac Colombia, caracterizado por la tendencia al alza y fuertemente por la estacionalidad anual de sus ventas.

Se observa como ambos modelos logran capturar de buena manera su estacionalidad anual, la cual se destaca en incremento de ventas para meses como marzo, agosto o noviembre y bajas considerables para meses como abril.

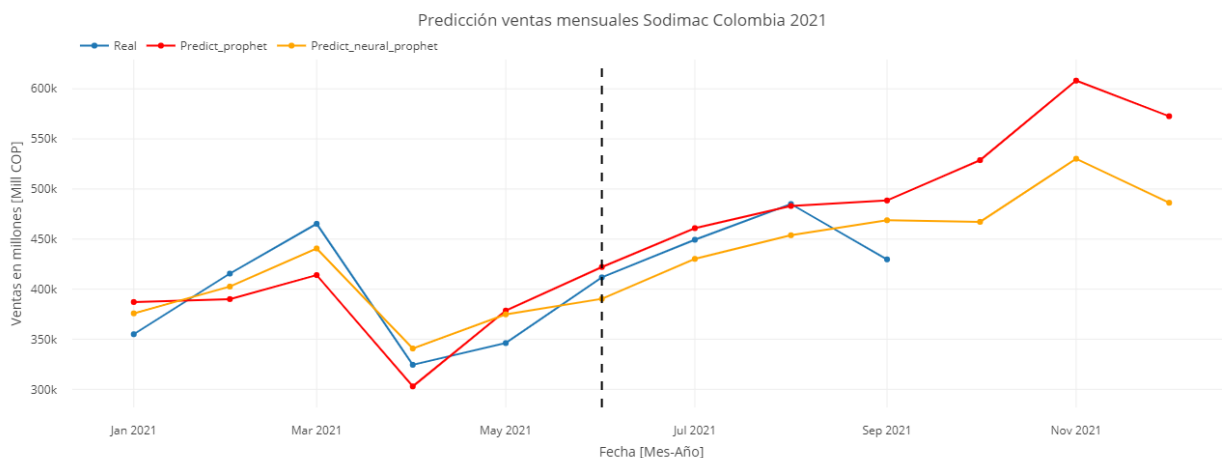


Figura 4.17: Predicción ventas mensuales - Sodimac Colombia

De esta manera ambos modelos proyectan ventas para el último trimestre de 2021 en torno a los 500.000 millones de pesos colombianos [COL], donde Prophet se acerca de mejor manera a la predicción de julio-septiembre que la proyección de NeuralProphet, sin embargo, tiende a proyectar en mayor medida las ventas para fin de año.

A partir de las métricas de error de la tabla 4.13, se contraste este rendimiento de las predicciones de los modelos, donde Prophet posee un menor error absoluto (MAE %) de 5.77 % versus el 6.55 % de MAE % de NeuralProphet, 0.78 % mayor.

Tabla 4.13: Métricas predicción ventas mensuales Colombia

Métricas Mensual Colombia	Sesgo	MAE	MAE %	RMSE	RMSE %
Prophet	26219.55	26219.55	5.77 %	36957.43	8.13 %
Neural Prophet	1668.36	30839.20	6.55 %	33442.24	7.36 %

Sin embargo, analizando el sesgo de predicción y la raíz del error cuadrático (RMSE), vemos como NeuralProphet obtiene mejores métricas, con énfasis en el error promedio (sesgo) de 1668.36 [Mill COP] el cual es casi 16 veces menor a los 26219.55 [Mill COP] de Prophet. Esto se explica principalmente a que Prophet siempre sobrestima las ventas mensuales, en comparación a NeuralProphet que para julio y agosto genera predicciones menores a la venta real de esos meses.

A partir de estas métricas, se decide escoger los resultados de NeuralProphet al considerar que se ajusta de mejor manera a la tendencia para los últimos meses del año y en especial por la disminución de ventas de septiembre que se captura de mejor manera por las predicciones de NeuralProphet.

Argentina

Se procede con la predicción de ventas para el último trimestre de 2021 en Sodimac Argentina, las cuales llegan a niveles sobre los 1200 millones de pesos argentinos [ARS] tal como se observa en la figura 4.18.

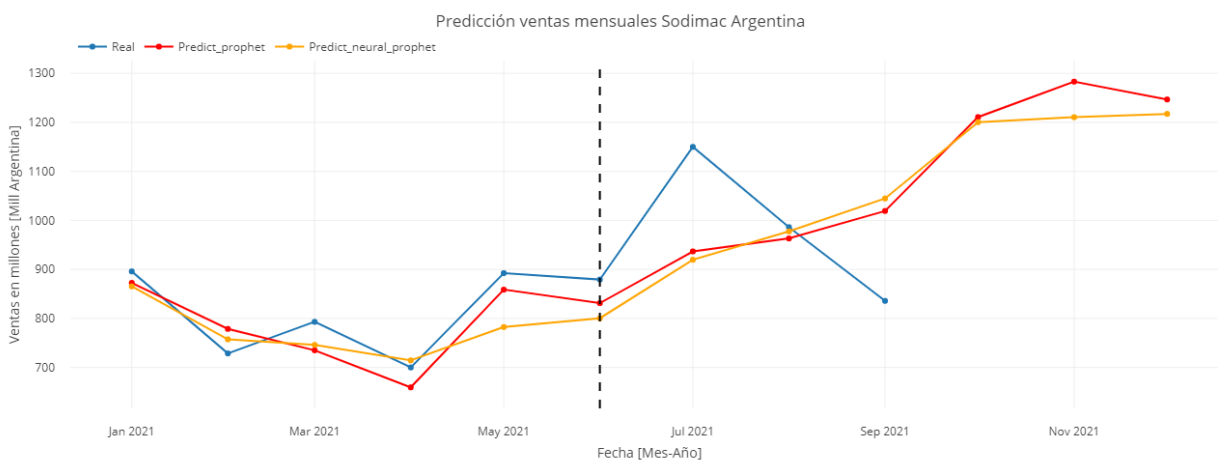


Figura 4.18: Predicción ventas mensuales - Sodimac Argentina

Se observa como ambos modelos logran captar la tendencia al alza de los últimos meses y peaks característicos de la estacionalidad de ventas mensuales en Argentina, donde tienen a disminuir el primer trimestre del año para recuperarse en los últimos dos.

Un aspecto a notar, es que las ventas de septiembre presentan ventas menores a años anteriores donde se caracterizara por ir constantemente al alza, y esto genera que ambos modelos no logren predecir las ventas de para aquel mes de buena manera.

Esto se refleja en las métricas de error de las predicciones de ambos modelos presentes en la tabla 4.14, las cuales llegan 14.13 % y 16.96 % de MAE % para Prophet y NeuralProphet, ambas claramente sobre los dígitos de error.

Tabla 4.14: Métricas predicción ventas mensuales Argentina

Métricas Mensual Argentina	Sesgo	MAE	MAE %	RMSE	RMSE %
Prophet	-17.62	139.96	14.13 %	163	16.45 %
Neural Prophet	15.62	168.03	16.96 %	199.18	20.11 %

En general, tanto para el MAE como el RMSE, sus equivalentes proporcionales y el error promedio, ambos algoritmos obtienen resultados aceptables pero no los mejores en comparación a otros países. De igual manera, comparativamente Prophet logra mejores métricas de error y visualmente logra capturar de mejor manera la estacionalidad anual de las ventas por lo cual es seleccionado para las predicciones finales de venta mensual en Sodimac Argentina.

Brasil

En la figura 4.19 se logra visualizar la predicción de ventas mensuales para Sodimac Brasil, a partir de Prophet como NeuralProphet. Se observa como ambos modelos logran capturar la estacionalidad anual de las ventas en Brasil, con disminución de ventas en marzo de cada año para ir recuperándose en los meses de junio y octubre.

De igual manera, para el último trimestre de 2021 se observa como ambos modelos pronostican ventas del orden de los 155 millones de reales brasileños [BRL], donde NeuralProphet tiende a pronosticar en todos los meses montos ligeramente mayores de venta llegando a los 160 [Mill BRL] para octubre y noviembre de 2021.

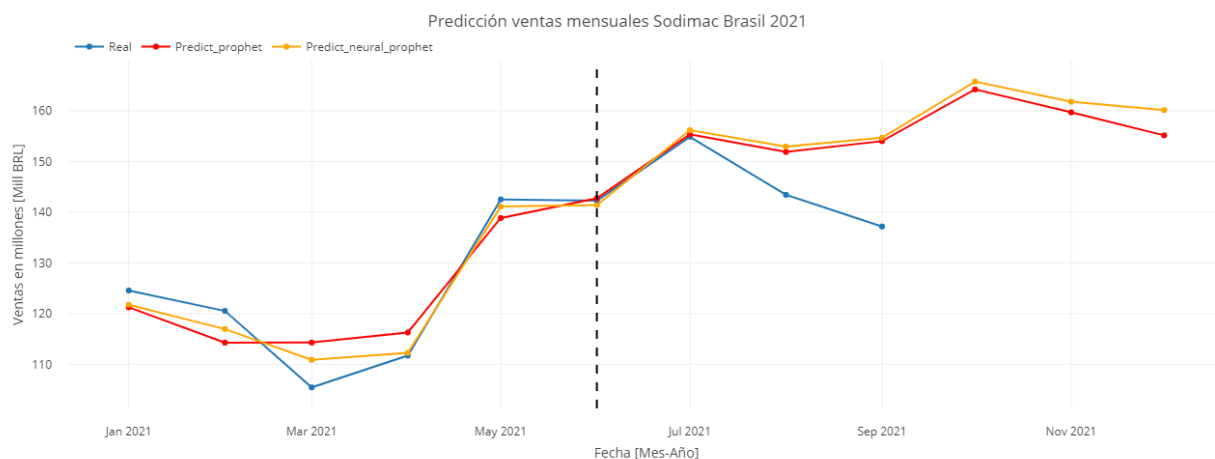


Figura 4.19: Predicción ventas mensuales - Sodimac Brasil

Estas predicciones, al igual que en Argentina, se alejan del valor real de ventas en septiembre del presente año frente a su notable caída de ventas y cuyo efecto no es perceptible en la historia pasada cercana. Esta diferencia impacta en las métricas de error disponibles en la tabla 4.15 y en especial en el sesgo positivo de ambos modelos.

Tabla 4.15: Métricas predicción ventas mensuales Brasil

Métricas Mensual Brasil	Sesgo	MAE	MAE %	RMSE	RMSE %
Prophet	8.59	8.59	5.92 %	10.88	7.49 %
Neural Prophet	9.45	9.45	6.51 %	11.53	7.95 %

De igual manera, pese a la diferencia de septiembre, se aprecia como ambos modelos poseen métricas de error menor al 10 % tanto para el error absoluto MAE % y la raíz del error cuadrático medio RMSE %. En general, Prophet logra mejores métricas, con un sesgo y MAE de 8.59 y para el MAE % de 5.92, 0.57 % menor al de NeuralProphet.

Uruguay

En la figura 4.20 es posible observar las predicciones de ventas para fines de 2021 de los modelos de machine learning implementados en Sodimac: Prophet y NeuralProphet.

En estas predicción se logra capturar la tendencia al alza de las ventas, sobre todo en el último año, con peaks propios de la estacionalidad anual en meses como marzo y junio. Visualmente se logra apreciar como el modelo de NeuralProphet obtiene mejor desempeño dado que se acerca más a las ventas reales de Sodimac Uruguay, llegando a predecir para el último trimestre ventas sobre los 280 [Mill UYU] e incluso el peak máximo de diciembre de 344 [Mill UYU].

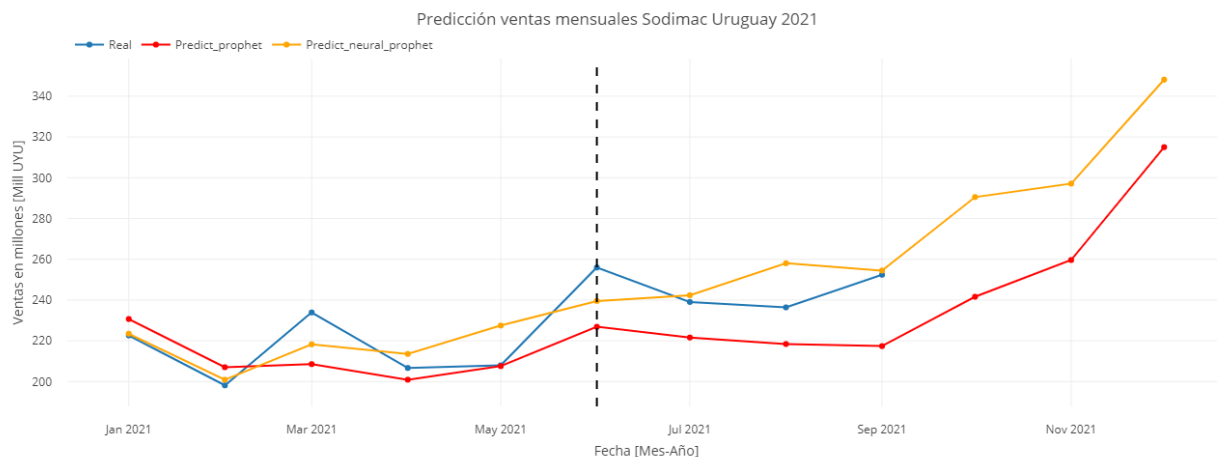


Figura 4.20: Predicción ventas mensuales - Sodimac Uruguay

Analizando las métricas de error de ambos modelos, disponibles en la tabla 4.16, se reafirma la perspectiva visual de desempeño de los modelos donde las predicciones basadas en NeuralProphet poseen un error absoluto MAE % de 3.45 %, casi 3 veces menor al MAE % de Prophet que llega a los 9.68 %.

Tabla 4.16: Métricas predicción ventas mensuales Uruguay

Métricas Mensual Uruguay	Sesgo	MAE	MAE %	RMSE	RMSE %
Prophet	-23.48	23.48	9.68 %	24.85	10.24 %
Neural Prophet	8.38	8.38	3.45 %	12.41	5.12 %

Otro aspecto a notar es el sesgo promedio de las predicciones, donde el modelo Prophet tiende a subestimar las ventas a diferencia de los resultados de NeuralProphet que sobrestiman las ventas.

De igual manera, el mejor desempeño de las predicciones basadas en el modelo de NeuralProphet son notoriamente mejores y se escogen para la comparación final de ventas mensuales de Sodimac Latam.

México

Se procede con los resultados de predicción de ventas para fines de 2021 en Sodimac México que se visualiza en la figura 4.21, caracterizadas por la fuerte estacionalidad anual de sus ventas con peaks en meses de marzo, mayo o julio y su disminución consecuentes en las ventas de meses como febrero, abril o junio.

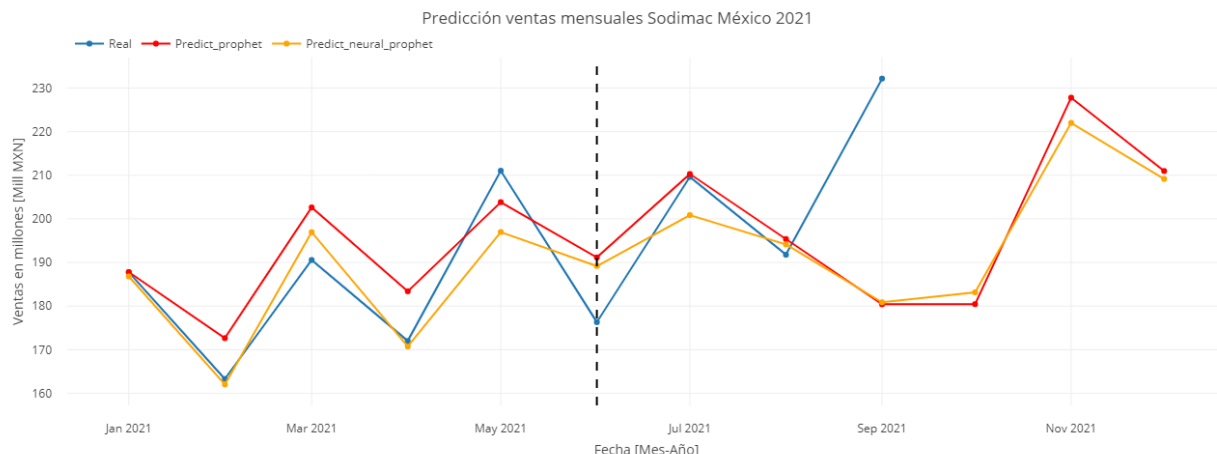


Figura 4.21: Predicción ventas mensuales - Sodimac México

Desde las predicciones de ventas de ambos modelos, roja para Prophet y naranja para NeuralProphet, se observa como ambas captan correctamente la tendencia al alza de las venta y la estacionalidad en forma de sierra que se describía anteriormente. Se logra apreciar que claramente la curva roja del modelo Prophet logra acercarse en mayor medida a las ventas reales para el 2021 y llegando a predecir para el último trimestre del año ventas sobre los 200 millones de pesos mexicanos [MXN].

Al igual que en modelo diario de ventas para México, la apertura de un nuevo local en el país genera un incremento de venta en septiembre que es resulta muy difícil de predecir debido a que no existe historia de las ventas de dicha sucursal de Sodimac. De igual manera, ambos modelos logran métricas menores al 10 % en el MAE %, tal como se aprecia en la tabla 4.17.

Tabla 4.17: Métricas predicción ventas mensuales México

Métricas Mensual México	Sesgo	MAE	MAE %	RMSE	RMSE %
Prophet	-15.82	18.66	8.84 %	29.93	14.17 %
Neural Prophet	-19,22	20.79	9.84 %	30.06	14.23 %

De esta, se confirma cuantitativamente el mejor desempeño de las predicciones en base a Prophet, con un MAE % de 8.84 % que es menor en 1 % al MAE % de NeuralProphet. En cuanto al RMSE %, ambos modelos poseen esta métrica en torno al 14 % y se observa como Prophet posee -3.4 mill menos de error promedio (sesgo) en su predicción.

Perú

Finalmente se muestran los resultados para las predicciones de venta mensual de Sodimac Perú en la figura 4.22 caracterizadas por un fuerte incremento de ventas para el último trimestre del año por su tendencia, en especial tras 2020 que afectó considerablemente las ventas por el Covid19.

Se logra apreciar como las ventas van en aumento desde los 300 [Mill PEN] al orden de los 500 [Mill PEN] para el modelo de Prophet y 400 [Mill PEN] para los resultado del modelo NeuralProphet.

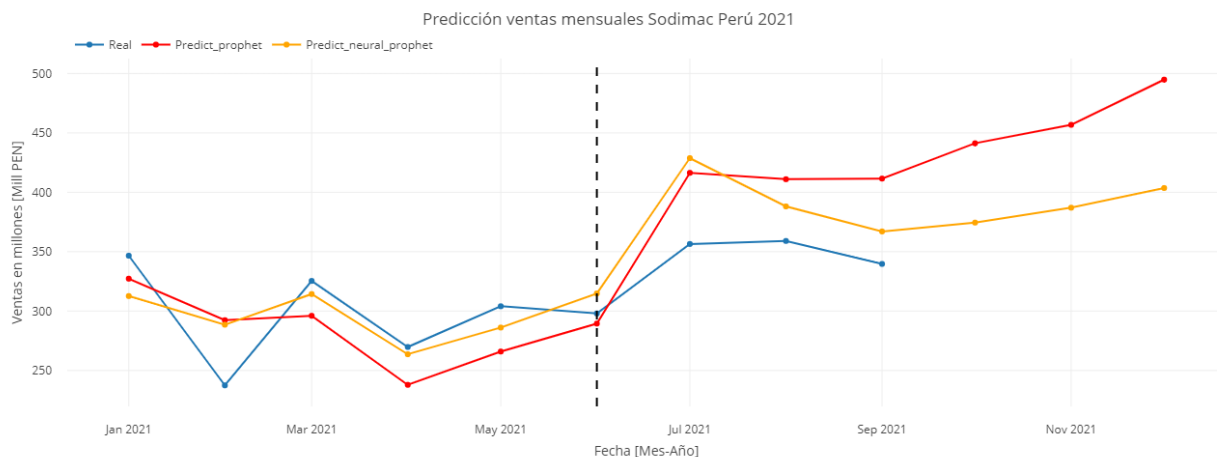


Figura 4.22: Predicción ventas mensuales - Sodimac Perú

Visualmente, tanto en los meses de entrenamiento (visibles previo a la vertical negra segmentada) como los meses del segundo semestre de 2021, el modelo en base a NeuralProphet logra acercarse de mejor manera a las ventas reales de Sodimac Perú, en especial para los meses de agosto y septiembre donde el modelo Prophet se aleja considerablemente en la proyección de ventas.

Este suceso se contrasta con las métricas de validación de la tabla 4.18, donde se aprecia como NeuralProphet logra mejor desempeño en todas las métricas de error.

Tabla 4.18: Métricas predicción ventas mensuales Perú

Métricas Mensual Perú	Sesgo	MAE	MAE %	RMSE	RMSE %
Prophet	61.25	61.25	17.41 %	61.79	17.57 %
Neural Prophet	34.69	34.69	9.86 %	41.47	11.79 %

En el detalle, se observa como los resultados de NeuralProphet presentan un error promedio (sesgo) positivo de 34.69 mill, lo cual refleja como el modelo sobrestima las ventas reales para julio, agosto y septiembre, pero en menor medida a los 61.25 mill que sobrestima el modelo Prophet.

De igual manera para el error absoluto porcentual MAE %, NeuralProphet predice con un 9.68 % de error en contraparte de los 17.41 % de error de los resultados de Prophet, 7.55 % menos de error que se refleja claramente en la visualización de las predicciones.

En base a lo anterior se selecciona los resultados del modelo de NeuralProphet para la comparación final de predicciones mensuales en Sodimac Latam.

Compilado de predicción de ventas mensuales Sodimac Latam

A partir de los resultados de predicción de ventas de los mejores modelos en cada país, se procede a compilar estos para visualizar netamente las predicciones de venta en todo el negocio de Sodimac.

En la figura 4.23 se observan las proyecciones de venta mensual para Colombia y Chile, ambas claramente superiores al resto de países de Sodimac y que llegan a ventas para los meses de octubre, noviembre y diciembre en promedio de 502.000 [Mill COP] y 275.000 [Mill CLP] respectivamente.

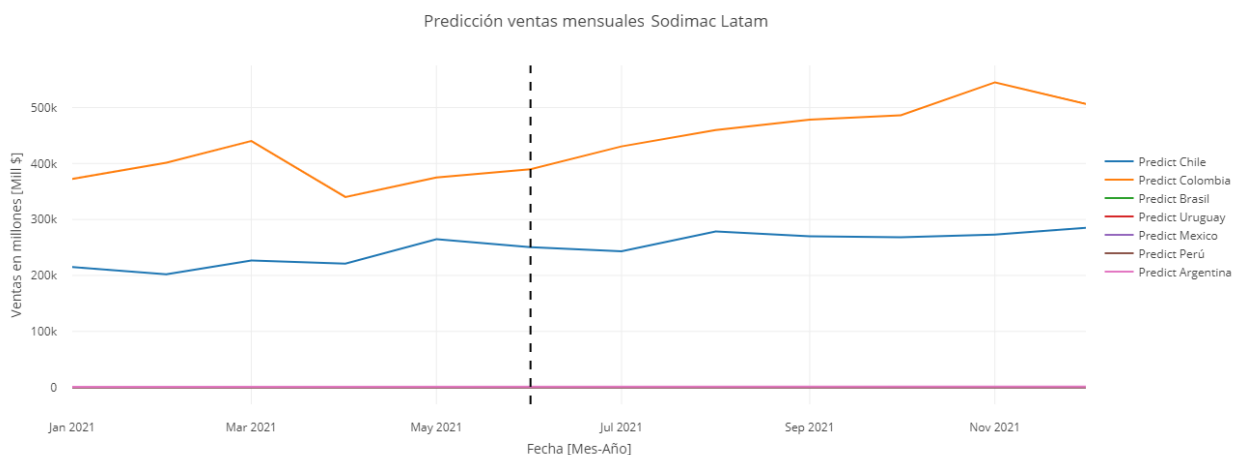


Figura 4.23: Predicción ventas mensuales Sodimac Latam 1

Filtrando las ventas de Chile y Colombia que disminuyen el detalle del resto de países, se logra observar el comportamiento de ventas a futuro de Sodimac Argentina en la figura 4.24, donde se predicen ventas en promedio para los últimos 3 meses del año de 1246 [Mill ARS].

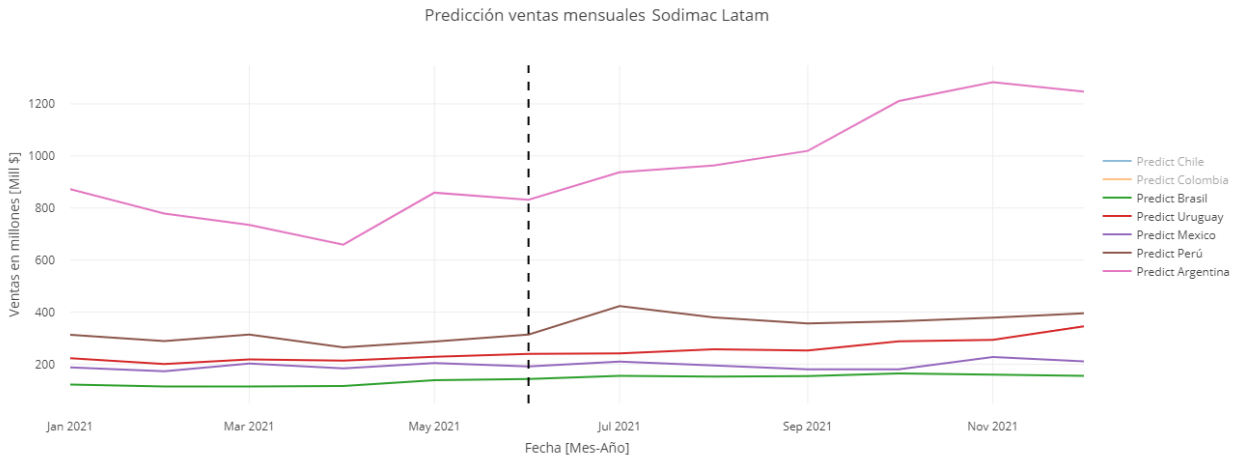


Figura 4.24: Predicción ventas mensuales Sodimac Latam 2

Finalmente, se muestra en la figura 4.25 la proyección final de ventas para los países con ventas mensuales menores a los 400 millones de su divisa interna: Brasil con la curva verde, Uruguay destacado con la curva roja, México con la línea morada y Perú el trazo café. Se aprecia como la tendencia al alza es propia para cada país, donde es más marcada para países como Perú y Uruguay.

Cuantitativamente se predicen ventas promedio para los meses de octubre, noviembre y diciembre de 159 [Mill BRL] para Brasil, 307 [Mill UYU] para Uruguay, 206 [Mill MXN] para México y 379 [Mill PEN] en Sodimac Perú respectivamente.

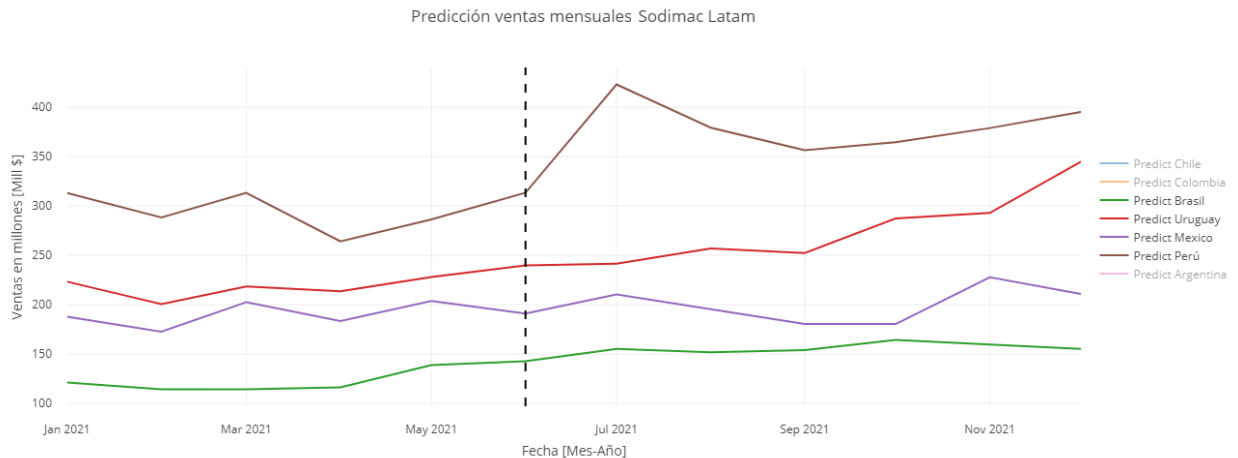


Figura 4.25: Predicción ventas mensuales Sodimac Latam 3

4.2.3. Impactos y beneficios del proyecto

A partir del trabajo desarrollado se logró diseñar e implementar el proyecto de predicción de ventas dentro de Sodimac Latam, en un área que ya sea por recursos, horas hombre u

otros aspectos, no realizaba un proceso de predicción de sus ventas. Este desarrollo permite generar predicciones precisas para resolución diaria y mensual basadas netamente en datos transaccionales de ventas y sus características a lo largo de la historia.

A su vez, a partir de las herramientas y tecnologías digitales con el cual se trabaja el proyecto se logra aprovechar sus propios beneficios y potencialidades. Desde los algoritmos de machine learning modernos que genera predicciones rápidas, interpretables y fácilmente ajustables; el trabajo mediante software computacional como Python que incorpora gran cantidad de librerías para el manejo de datos de manera rápida; y la utilización de servicios de la nube como bigquery de GCP que permite el almacenamiento de gran cantidad de datos sin limitarse por la capacidad local de los computadores.

A continuación se detallan 3 beneficios e impactos relevantes del trabajo en Sodimac: la precisión de sus resultados, la automatización hasta ahora y potencial automatización completa del proceso a futuro y la arista de accionables que se abre al poseer predicciones de venta a futuro y como impacta el trabajo en la unidad de negocio.

Precisión del proceso

Un beneficio propio de las predicciones en base a los modelos de machine learning se destaca en la precisión de estas en relación a las ventas reales del tercer trimestre del 2021, donde obtienen errores absolutos porcentuales MAE % menores al 15 % para la venta diaria y menor al 10 % para la venta mensual en mayor parte de los países, a excepción de Argentina, que por características propias de sus ventas posee errores absolutos % menores al 20 % en la resolución mensual mensual y 15 % en la diaria.

Estos resultados, pese a tener aun más campo de mejora, resultan errores aceptables en rubros característicos de venta minorista y sirven como un avance considerable en un área donde no se lograban realizar predicciones de venta actualmente.

Automatización del proceso

Mediante la utilización de programación en Python, librerías eficaces en el manejo de datos disponibles en este y la base de datos de ventas validada desde el área, se permite automatizar gran parte del proceso de predicción de ventas sin mayor interacción que ejecutar códigos de programación.

Esta automatización resulta de gran relevancia en el área, dado que una de las razones que incidían en no poseer un modelo de predicción de ventas en finanzas de Sodimac era la poca cantidad de horas hombre disponible para este tipo de tareas. De esta manera se generan activos analíticos que mediante programación realizan la predicción de ventas diarias y mensuales en no más de 10 minutos con sus entregables para el área.

Este proceso aún tiene más campo de automatización, dado que la base de datos actual que usa el modelo se obtiene desde el mismo área. Sin embargo, existe en desarrollo una base de datos en bigquery de Google Cloud Plataform que dispone toda la información de ventas para Sodimac, la cual cuando sea validada podrá reemplazarse como entrada a los modelos

de predicción de venta y de esta manera automatizar completamente el proceso.

Accionables del proceso

Un último impacto propio de desarrollar un proyecto en un área donde no realizaban actualmente predicciones de venta diaria o mensual es la capacidad de generar acciones financieras, comerciales o cualquier otra a partir de las mismas predicciones de venta.

La presencia de este activo predictivo logra generar información valiosa para el negocio la cual no se poseía anteriormente, permitiendo focalizar trabajo dentro del área con los elementos propios de la analítica avanzada: desde entender el comportamiento de las ventas, predecir las ventas a futuro (modelos implementados) a generar analítica prescriptiva y analizar que acciones se pueden realizar para obtener determinados resultados de venta a futuro.

4.3. Predicción de deuda acumulada para Falabella Inmobiliaria

En la presente sección se procede con los resultados para el proyecto de predicción de ventas mensuales de deuda acumulada en Falabella Inmobiliaria, unidad de negocio en el rubro de arriendo de locales, tiendas, centros comerciales, espacios publicitarios, entre otros.

Estos resultados se presentan en 3 niveles de apertura: para la deuda total del negocio, aperturado por las dos sociedades internas que posee el mismo negocio y finalmente las predicciones cuantitativas de deuda separada por tramos de antigüedad para el cálculo de provisiones de incobrables de la compañía.

Para estos 3 niveles, se exponen los resultados tanto del modelo Prophet como NeuralProphet, algoritmos de machine learning implementados para las predicciones de deuda acumulada donde se analizan sus resultados en base a las métricas de error y determinando el mejor modelo con su predicción respectiva.

4.3.1. Pronóstico de deuda acumulada del negocio

En la figura 4.26 se presenta los resultados de predicción de deuda acumulada mensual para el negocio de Falabella Inmobiliaria, con curva roja los resultados para Prophet y curva amarilla los resultados para NeuralProphet.

Ambos modelos logran predecir la deuda acumulada para fines de 2021 y todo 2022, sin embargo, a simple vista se logra observar un comportamiento desigual de sus predicciones donde Prophet proyecta una deuda a cierre de 2022 de 3021 [Mill CLP] mientras que para NeuralProphet llega a los 404 [Mill CLP].

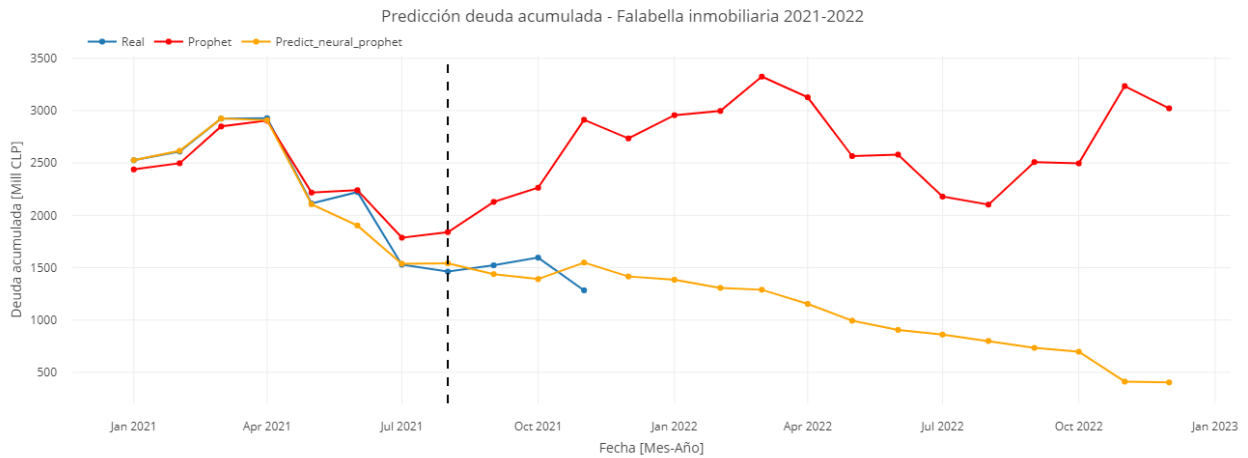


Figura 4.26: Predicción deuda acumulada mensual para todo el negocio de Falabella Inmobiliaria

En estos resultado dispar tiene su mayor explicación en como se adaptan los modelos a el cambio de tendencia y estacionalidad que se provoca por la bajada brusca de deuda acumulada que ocurre desde el segundo trimestre de 2021 (abril-mayo), donde como se vió en la caracterización de la data de deuda, tras llegar a su peak 2927 [Mill CLP] en 7 meses disminuye a 1283 [Mill CLP], una reducción cercana a los 1600 [Mill CLP] que impacta en la tendencia de ventas y de la cual no se posee precedentes en años anteriores para meses como agosto y noviembre donde solía subir la deuda acumulada, afectando la estacionalidad anual de ventas.

A partir de lo anterior, se observa como el modelo en base a Prophet no logra capturar de manera efectiva el cambio de tendencia los últimos meses y mantiene su tendencia semi estacionaria de deuda, en cambio el modelo de NeuralProphet con el ajuste de parámetros logra capturar de mejor manera este cambio de deuda y replica esta a 2022 bajando de los 1000 [Mill CLP] para mayo del 2020.

Esta apreciación visual se replica en las métricas de error del modelo de predicción para los meses de septiembre, octubre y noviembre de 2021, donde los resultados de deuda para NeuralProphet obtienen un 12.60% de error absoluto (MAE), un 13.61% de RMSE y un error promedio de -7.85 [Mill], muy pequeño en comparación a los resultados de Prophet.

Tabla 4.19: Métricas predicción deuda acumulada total negocio F. Inmobiliaria

Métricas total negocio	Sesgo	MAE	MAE %	RMSE	RMSE %
Prophet	967.58	967.58	65.95 %	1075.1	73.28 %
Neural Prophet	-7.85	184.87	12.60 %	199.65	13.61 %

Estos resultados, considerando que el cambio de comportamiento de la deuda acumulada se origina principalmente por factores externos como gestión de cobranza desde Falabella Inmobiliaria resultan un correcta aproximación al futuro, entendiendo que la reducción de

deuda acumulada de la compañía hace sentido con la realidad del negocio, puesto que no es sano financieramente poseer tanta deuda impaga por sus clientes.

A partir de los resultados y las métricas de error explayadas en la tabla 4.19, se determina los resultados de NeuralProphet como los finales para la predicción de deuda acumulada del negocio. En la figura 4.27 se refleja esta predicción final, donde la deuda acumulada real de proyecta a llegar bajo los 500 millones de pesos para noviembre y diciembre de 2020.

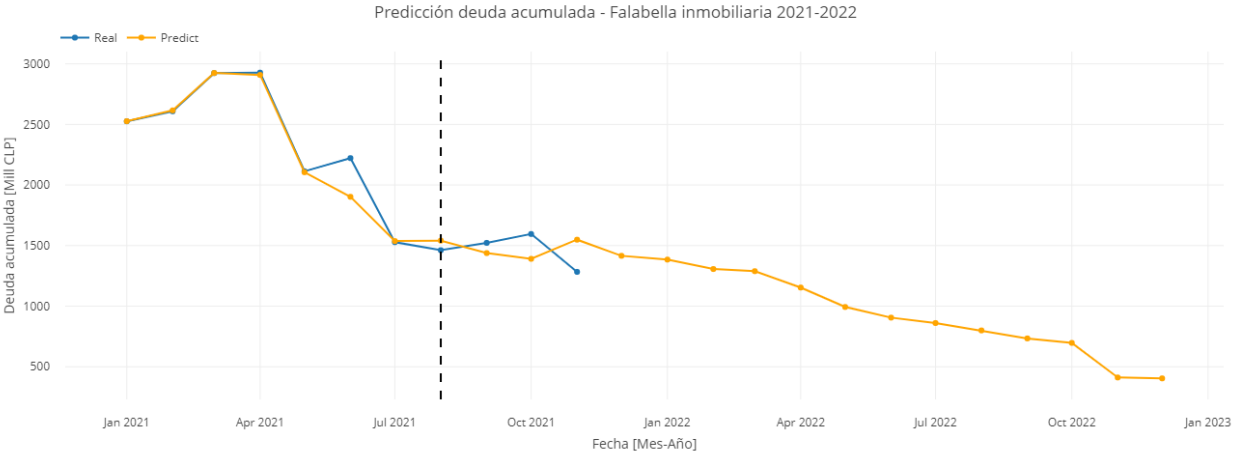


Figura 4.27: Resultados de predicción de deuda acumulada mensual para todo el negocio de Falabella Inmobiliaria

4.3.2. Pronóstico de deuda acumulada por sociedades

Se procede con las predicciones de deuda acumulada para las sociedades internas de Falabella Inmobiliaria, propias del negocio de arriendo de locales y lugares físicos en las propiedades de la unidad de negocio.

Sociedad C069

En la figura 4.28 se muestran los resultados de deuda acumulada proyectada para fines de 2021 y todo 2020, caracterizados los resultados de Prophet como NeuralProphet con la pendiente o tendencia negativa de la deuda para los próximos meses.

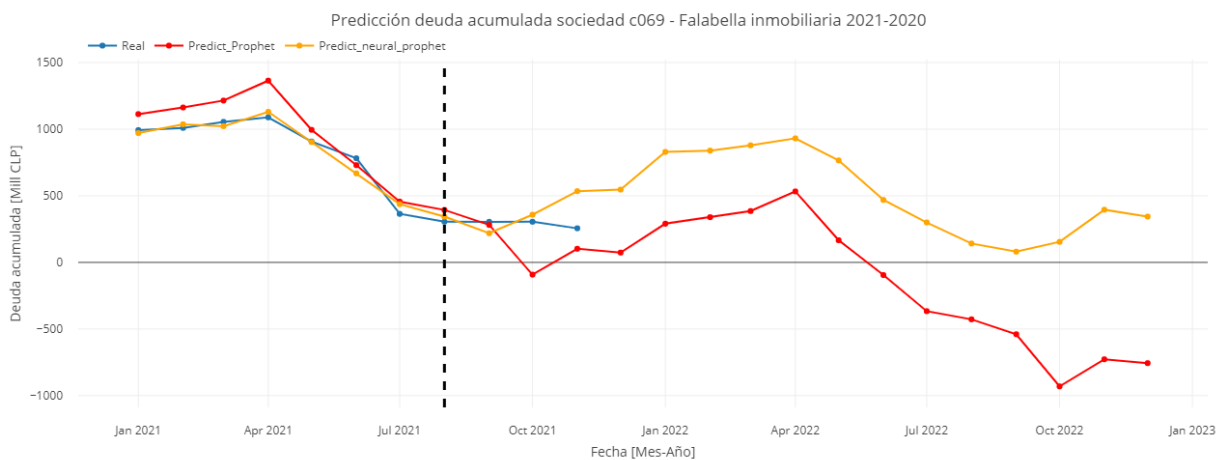


Figura 4.28: Predicción deuda acumulada mensual para sociedad C069 de Falabella Inmobiliaria

Esta sociedad se caracteriza por poseer gran parte de la disminución de deuda acumulada desde abril de 2021, pasando de 1088 [Mill CLP] a 255 [Mill CLP] en noviembre del mismo año, y en general, desde el análisis descriptivo de la deuda acumulada se observa como la sociedad C069 desde abril de 2020 posee una tendencia negativa.

Pese a que ambos modelos capturan esta tendencia a la baja, ocurre que el modelo en base a Prophet traslada este comportamiento a una deuda acumulada negativa, la cual en términos reales del negocio carece de sentido. En cambio, el modelo en base a NeuralProphet logra estabilizar su predicción de resultados y replicar fenómenos estacionales del pasado como incremento de deuda para inicios del 2022.

Esto se replica en las métricas de error de ambos modelos de la tabla 4.20, donde se observa como los resultados sobre los meses de septiembre, octubre y noviembre de 2021 del modelo en base a NeuralProphet logra obtener ampliamente mejor desempeño, sobre todo en el error absoluto de 48.26 % que es 18.01 % menor al generado por los resultados del modelo Prophet.

Tabla 4.20: Métricas predicción deuda acumulada Sociedad C069

Métricas C069	Sesgo	MAE	MAE %	RMSE	RMSE %
Prophet	-191.26	191.26	66.27 %	246.02	85.24 %
Neural Prophet	81.92	139.28	48.26 %	171.23	59.33 %

Sin embargo, se observa como las métricas de error para esta sociedad son altas en comparación a los otros modelos desarrollados, lo cual se explica por la diferencia de deuda acumulada real y proyectada para el mes de noviembre.

Lo anterior se entiende principalmente por los efectos de la gestión de cobranza en el corto plazo que impactaron en el comportamiento de la deuda acumulada desde mayo en adelante. Además se suma que la historia pasada para años como 2019 y 2020 posee características

propias del contexto del país como el fenómeno social de 2019 y la pandemia por el covid de 2020 que impactaron fuertemente el rubro de malls, tiendas comerciales y otros relacionados con el arriendo de locales y espacios en Falabella Inmobiliaria, traduciéndose en incrementos de deuda como el ocurrido en noviembre de 2019.

Estos aspectos abren la arista de potenciales mejoras a los modelos desarrollados a futuro, donde poder integrar elementos cuantitativos desde la gestión de cobranza o facturación/pago tras contexto del país permitiría disminuir el error de las predicciones de deuda a futuro.

De igual manera, los resultados del modelo NeuralProphet explayados en la figura 4.29 poseen sentido en su comportamiento macro y sirven como referencia para la deuda acumulada de la sociedad C069 a futuro donde se espera llegar a 344 millones de pesos.

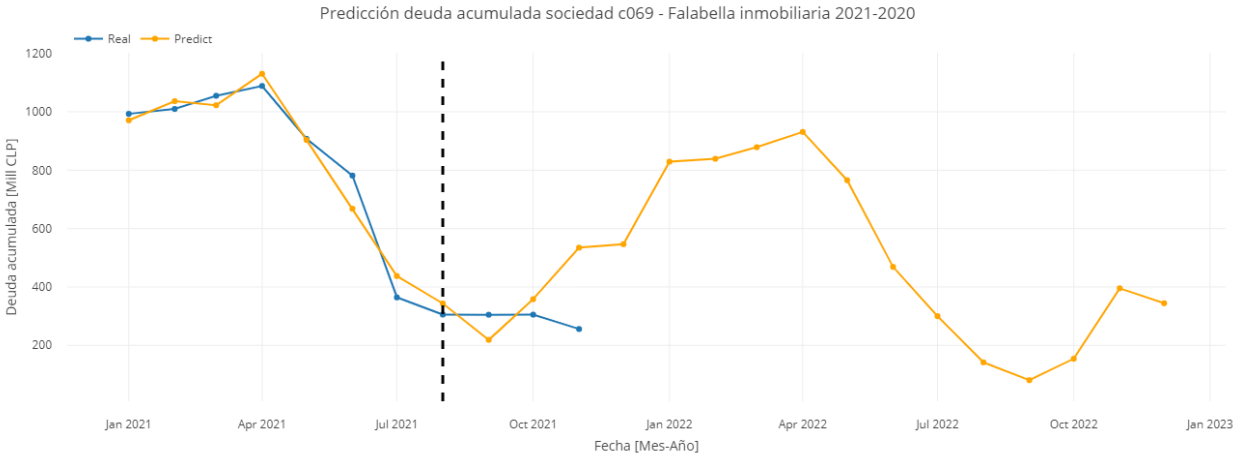


Figura 4.29: Resultados de predicción deuda acumulada mensual para sociedad C069 de Falabella Inmobiliaria

Sociedad C106

Se procede con los resultados de los modelos desarrollados para la deuda acumulada real de la sociedad C106 de Open Plaza, los cuales se ilustran en la figura 4.30.

En esta se logra apreciar, al igual que la predicción para venta de todo el negocio, como los modelos generan resultados dispares de acuerdo a como logran capturar el cambio de tendencia en los últimos meses del año, la cual pasa de un fuerte incremento en 2020 para estabilizarse a la baja desde mayo del 2021.

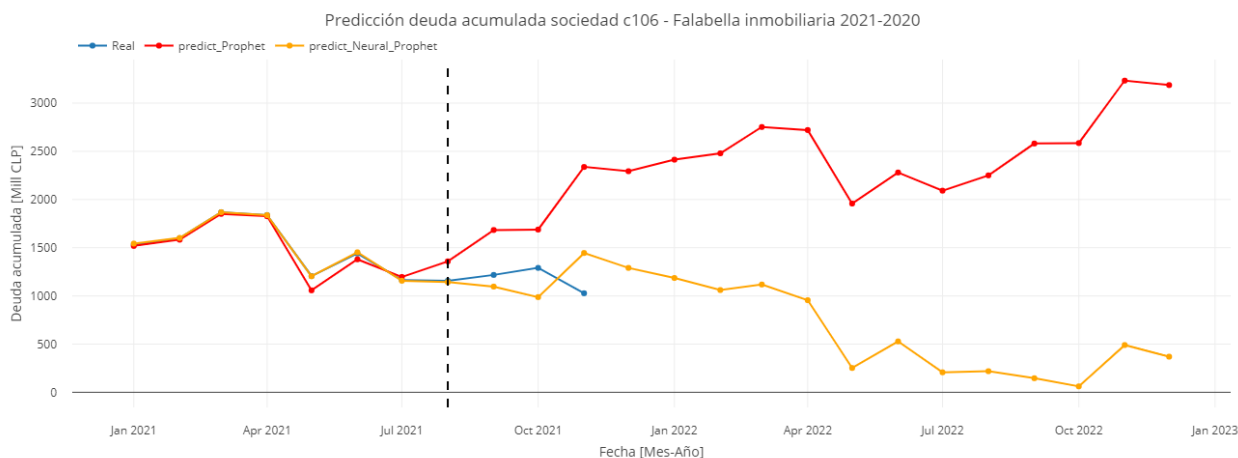


Figura 4.30: Predicción deuda acumulada mensual para sociedad C106 de Falabella Inmobiliaria

De esta manera, se observa como el modelo en base a Prophet mantiene la tendencia al alza para todo 2022 llegando cercana a los 3500 [Mill CLP], en cambio, los resultados para el modelo en base a NeuralProphet adquieren el cambio de tendencia a la baja desde fines de 2020, proyectando una deuda futura que logra bajar de los 1000 millones de pesos en 2022.

Estos resultados se condicen con las métricas de error de la tabla 4.21, obtenidas en comparación a la deuda real para septiembre, octubre y noviembre de 2021. Es esta se advierte como el error absoluto porcentual de 23.82 % de NeuralProphet es considerablemente menor al los resultados de Prophet que obtienen un 61.49 %, más del doble de error debido a que Prophet pronostica una deuda en aumento constante a futuro, la cual se aleja considerablemente a la estabilización y bajada que ocurre en noviembre de 2021.

Tabla 4.21: Métricas predicción deuda acumulada Sociedad C106

Métricas C106	Sesgo	MAE	MAE %	RMSE	RMSE %
Prophet	724.70	724.70	61.49 %	835.17	70.87 %
Neural Prophet	53.74	280.75	23.82 %	306.21	25.98 %

De esta manera, se definen los resultados para el modelo NeuralProphet como los resultados finales para la sociedad C106 de Falabella Inmobiliaria, los cuales se observan de mejor manera en la figura 4.31. Es estos se proyecta una deuda acumulada a la baja que llega a los 337 millones para diciembre de 2022, resultado que hace sentido con los esfuerzos de los últimos meses de disminuir su deuda acumulada.

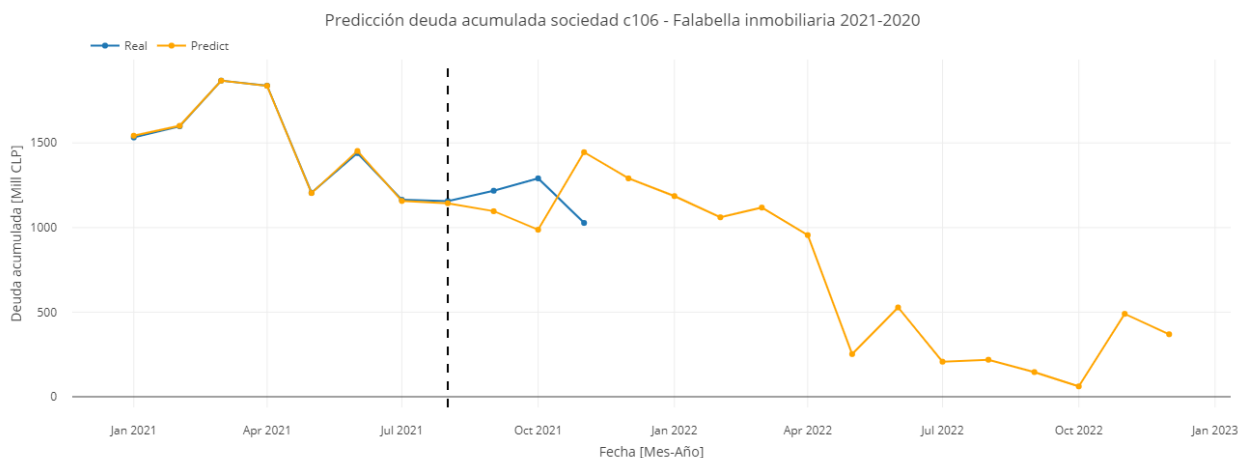


Figura 4.31: Resultados de predicción deuda acumulada mensual para sociedad C106 de Falabella Inmobiliaria

4.3.3. Impactos y beneficios del proyecto

El desarrollo de los modelos de predicción de deuda acumulada tanto para todo el negocio como para su apertura por sociedades logra cumplir correctamente el objetivo del proyecto y la necesidad de Falabella Inmobiliaria de modernizar sus procesos financieros, ya sea parcial o completamente.

De esta manera, el uso de tecnologías digitales, elementos desde la analítica avanzada y principalmente el uso de algoritmos de machine learning permite no tan solo generar predicciones precisas y rápidas de realizar, sino que, logra generar beneficios propios de estas tecnologías y extender usos a potenciales nuevos proyectos dentro de la unidad de negocio a futuro.

Metodología y automatización en limpieza de datos para modelos predictivos

Un elemento propio de la caracterización y preprocesamiento de los datos desde la base de datos de SaS fue el desarrollo de reglas, validaciones, y en general, una metodología para la transformación de la data bruta de SaS hacia la deuda acumulada real por mes para todo el negocio y sus sociedades C069 y C106.

Considerando que la base en SaS posee todo el flujo de documentos asociados al negocio, esta posee información en constante actualización, con documentos o “partidas” cerradas y otros no asociados a cuentas relacionadas con la deuda. Por esto, se definieron reglas y filtros validados para generar correctamente esta base con la deuda real acumulada por mes.

- Filtro de partidas abiertas al cierre de cada mes, lo que permite extraer correctamente la deuda acumulada desde SaS.
- Filtro de las 11 cuentas asociadas documentos de facturación, pago, notas de créditos, etc. que inciden en la deuda acumulada de Falabella Inmobiliaria.

- Filtro de días de mora mayores a cero para extraer correctamente la deuda real a cada mes.

Estas reglas, junto con su definición y validación desde el negocio, se transforman a fin de cuentas en una metodología para la extracción y consolidación de datos desde la base en SaS para la implementación de modelos de machine learning, tanto para la predicción de deuda como otros posibles a desarrollar.

Además, estas reglas y filtros se automatizan mediante programación en software como Python y R, aprovechando sus librerías útiles y eficaces en el manejo de datos. Esto resulta de gran relevancia en el área, dado que es un activo analítico para posibles otros proyectos y libera horas hombre en caso de realizar estos mismos filtros en excel de manera manual.

Posibilidad de expansión de modelo predictivo de deuda

Otro impacto propio del desarrollo de este tipo de modelos es la potencial expansión y aplicación para otras tareas dentro del área de finanzas de Falabella Inmobiliario.

Dentro de estos se destacan la posibilidad de aperturar aún más el detalle de deuda acumulada del negocio, esto a nivel de clientes de Falabella Inmobiliaria y sus sociedades. Esto permite, de acuerdo a su comportamiento de facturación, pago y deuda de los clientes, generar predicciones de deuda acumulada a futuro en base a los mismos algoritmos de machine learning utilizados hasta ahora o incluso agregar otros más.

Otra posible expansión del desarrollo realizado se enfoca en al predicción de deuda acumulada por tramos de antigüedad de la misma deuda, segmentando esta de acuerdo a tramos de días definidos en el negocio que se utilizan para el cálculo de provisiones para deuda incobrable por mes, proceso el cual tiene impacto en el presupuesto del negocio.

Estos ejemplos de expansión de los activos analíticos desarrollados permiten hacer un uso provechoso de los datos del negocio, fomentando las decisiones en base a datos (o “data-drive” del inglés) y, a fin de cuentas, generar valor y aprovechar los beneficios propios de las tecnologías en procesos financieros.

Capítulo 5

Conclusiones

En el presente trabajo, contextualizado en la transformación digital de una gran compañía como Falabella S.A., se logró diseñar e implementar 3 proyectos de transformación digital basados en analítica y algoritmos de machine learning que generaron mejoras en los procesos de predicción de venta para Tottus y Sodimac y en la predicción de deuda acumulada en Falabella Inmobiliaria, dando cumplimiento al objetivo principal de esta memoria.

Estos proyectos se basaron principalmente en aplicaciones desde la inteligencia artificial, algoritmos de machine learning enfocados en la predicción de series de tiempo como Prophet y NeuralProphet, la analítica con foco descriptivo y predictivo, y elementos propios de la transformación digital como la digitalización y automatización de procesos, conceptos los cuales fueron abarcados inicialmente en una revisión bibliográfica para avalar del trabajo desarrollado.

A partir de esta base teórica se generó una metodología de trabajo inspirada en metodologías desarrolladas para proyectos de ciencia de datos, la cual abarcó los objetivos específicos de la memoria: detectar necesidades reales en las unidades de negocio, desarrollar activos analíticos descriptivos y predictivos mediante el uso de los datos de las empresas, el uso de algoritmos de machine learning interpretables y vanguardistas como Prophet y NeuralProphet para predicciones de series de tiempo, y finalmente medir cuantitativa y cualitativamente las mejoras y beneficios en los procesos de las unidades de negocio.

En la práctica, mediante la implementación de esta metodología se logró abarcar en cada uno de los proyectos todas sus etapas de trabajo; desde la definición de su objetivo principal a solucionar mediante el uso de machine learning y analítica, el trabajo con datos desde su extracción a preprocesamiento y la justificación de selección de los algoritmos de Prophet y NeuralProphet en los tres proyectos realizados, esto sustentado que se buscó utilizar algoritmos vanguardistas, entendibles, interpretables en sus parámetros y que no funcionaran como caja negra, especialmente al considerar que se trabajó con usuarios no expertos en el área de ciencia de datos y machine learning. De esta manera, se logró obtener los resultados de cada uno de los proyectos basados en analítica predictiva y machine learning.

Para el primer proyecto en Tottus Chile, se logró implementar un modelo de predicción de ventas netas basado en Prophet con resultados de 108 [Mill CLP] de ventas para diciembre

de 2021 y con un error de predicción absoluto porcentual (MAE %) menor al 5 %, la cual significaba una mejora del 58 % en comparación al proceso actual de predicción de ventas existente en el área.

El segundo proyecto desarrollado logró la generación de predicciones de deuda mensual y diaria para las ventas de Sodimac en los 7 países que posee presencia en latinoamérica,. Estas predicciones se realizaron en base a algoritmos machine learning como Prophet y NeuralProphet, analizando sus desempeño en base a métricas de error y escogiendo la predicción de ventas del mejor modelo. De esta manera se logró predecir las ventas diarias y mensuales para cierre de 2021, con errores absolutos porcentuales (MAE %) menores al 20 % para la venta diaria y menores al 15 % para la venta mensual.

Para el tercer y último proyecto, se logró desarrollar un modelo predictivo de deuda acumulada para Falabella Inmobiliaria, generando predicciones mensuales de deuda que capturarán la dinámica del negocio y el comportamiento de pago de sus clientes. Estas predicciones de deuda acumulada obtuvieron errores absolutos porcentuales de 12.6 % para todo el negocio, 48.26 % para la sociedad C069 y 23.82 % para la sociedad C106.

Junto a los resultados de predicciones para los 3 proyectos de transformación digital, de manera general se lograron impactos y beneficios en los procesos financieros involucrados en esta incorporación de tecnologías digitales y aplicaciones del machine learning. Propio de los algoritmos predictivos se lograron pronósticos cuantitativos precisos, con errores en general menores al 15 % y que significaban una mejora frente a procesos actuales, e incluso, reactivando procesos en negocios como Sodimac.

Desde el uso de tecnologías y servicios de la nube se logró generar metodologías y automatizar gran parte de estos procesos, desde la extracción y manejo de datos hasta la aplicación y obtención de resultados de los mismos algoritmos de machine learning. Por último, una tercera pero no menos relevante arista de impactos, se destaca el potencial de expansión de los activos analíticos en otros procesos, realizando predicciones de ventas en aperturas de los datos que no se consideraban inicialmente e incluso en otras unidades de negocios de Falabella S.A., aportando con esto al trabajo y toma de decisiones en base a datos de los negocios.

Como conclusión general, en un mundo empresarial donde las iniciativas de adopción de tecnologías y la necesidad de adaptarse a un mundo no presencial se vieron acrecentadas con la pandemia, se logró apreciar como este tipo de proyectos de transformación digital reflejan un claro ejemplo donde resultados analíticos y comparables generan mejoras procesos financieros y aportan en la transformación digital de las organizaciones, en este caso, para la gerencia de finanzas de Falabella S.A.

5.1. Trabajo Futuro

A partir del trabajo desarrollado, se han identificado distintas oportunidades de mejora en los proyectos, en especial para los algoritmos de predicción implementados. Algunas de estas son:

- Integrar covariables o regresores que incorporen información propia del rubro del negocio como calendarios comerciales en el caso de la venta minorista de Tottus y Sodimac, o información de gestión de cobranza para el estudio de la deuda de Falabella Inmobiliaria.
- Probar en mayor amplitud configuraciones de parámetros en los modelos, en especial para el caso de NeuralProphet y su componente de la AR-NET.

En aspectos al desarrollo mismo de proyectos en las unidades de negocio se plantea:

- Integrar los algoritmos predictivos utilizados, e incluso nuevos algoritmos, en modelos productivos para la realización de predicciones conocidos por auto-ml, donde de acuerdo a los datos, su resolución y horizonte de pronóstico, se prueben distintos algoritmos de machine learning para la predicción objetivo y se entregue resultados del mejor algoritmo de manera automática.
- Dentro de los alcances de este trabajo se enfocó en modelos predictivos de series de tiempo, sin embargo, el dominio de información en las unidades de negocio permite explorar el uso de otros tipos de algoritmos de machine learning como clasificaciones, segmentaciones y cualquier otra que permitiese innovar y generar beneficios en los procesos de las gerencias de finanzas de Falabella S.A.

Bibliografía

- [1] P. C. Verhoef, T. Broekhuizen, Y. Bart, A. Bhattacharya, J. Q. Dong, N. Fabian, and M. Haenlein, “Digital transformation: A multidisciplinary reflection and research agenda,” *Journal of Business Research*, vol. 122, pp. 889–901, 2021.
- [2] AIMultiple, “85+ digital transformation stats from reputable sources [2021],” 2020.
- [3] K. Chandra, F. Plaschke, and I. Seth, “Memo to the cfo: Get in front of digital finance—or get left back,” 2018.
- [4] IFAC, “Case study: Maersk - future of the cfo and finance function,” 2020.
- [5] PricewaterhouseCoopers, “How a digital finance transformation simplified a complex industrial giant: Pwc,” 2019.
- [6] Deloitte, “Finance digital transformation case studies,” 2018.
- [7] I. C. Education, “What is robotic process automation (rpa)?,” 2020.
- [8] AppZen, “Winning at finance transformation 2 contents,” 2021.
- [9] S. Vennam, “What is cloud computing?,” 2020.
- [10] V. Kotu and B. Deshpande, *Data Science : Concepts and Practice*. Morgan Kaufmann Publishers, 2019.
- [11] D. Analytics, “Data analytics - models and algorithms for intelligent data analysis | thomas a. runkler | springer,” 2020.
- [12] J. Saltz, I. Shamsurhin, and K. Crowston, “Comparing data science project management methodologies via a controlled experiment.”
- [13] R. Timaran, I. H. Arteaga, J. Caicedo, A. H. Troya, and J. Carlos, *Descubrimiento De Patrones De Desempeño Académico Con Arboles De Decisión En Las Competencias Genéricas De La Formación Profesional*. Universidad Cooperativa De Colombia, 2016.
- [14] U. Fayyad and R. Uthurusamy, “Data mining and knowledge discovery in databases,” *Communications of the ACM*, vol. 39, pp. 24–26, 1996.
- [15] SAS, “Sas enterprise miner 15.2: Reference help,” 2021.
- [16] P. Chapman, J. Clinton, R. Kerber, T. Reinartz, C. Shearer, and R. Wirth, *CRISP-DM 1.0: step-by-step Data Mining Guide*. SPSS, 2000.
- [17] S. J. Russell and P. Norvig, *Artificial intelligence : a modern approach*. Prentice-Hall, 2010.
- [18] I. C. Education, “What is artificial intelligence (ai)?,” 2020.

- [19] F. Ruehle, “Data science applications to string theory,” *Physics Reports*, vol. 839, pp. 1–117, 2020.
- [20] I. C. Education, “What is unsupervised learning?,” 2020.
- [21] I. C. Education, “What are neural networks?,” 2020.
- [22] A. Geron, *Hands-on Machine Learning with Scikit-Learn and TensorFlow concepts, tools, and Techniques to Build Intelligent Systems*. O’Reilly Media, Inc., 2019.
- [23] R. J. Hyndman and G. Athanasopoulos, *Forecasting : Principles and Practice*. Otexts, 3er edition ed., 2021.
- [24] S. J. Taylor and B. Letham, “Forecasting at scale,” 2017.
- [25] O. Triebe, H. Hewamalage, P. Pilyugina, N. Laptev, C. Bergmeir, and R. Rajagopal, “Neuralprophet: Explainable forecasting at scale,” 2021.
- [26] O. Triebe, N. Laptev, and R. Rajagopal, “Ar-net: A simple auto-regressive neural network for time-series a preprint,” 2019.
- [27] N. Vandepuit and W. De, *Data Science for Supply Chain Forecasting*. Berlin De Gruyter, 2019.