



**Modelo de Machine Learning para la eficiencia operativa del portafolio de productos
y servicios del área de Cultura, Recreación y deporte de una Caja de Compensación en
Colombia**

**GIOVANNY ROBERTO GONZÁLEZ PEDRAZA
JUAN CAMILO CIFUENTES RUIZ**

Universidad EAN
Facultad de Ingeniería
Maestría en Inteligencia de Negocios
Bogotá, Colombia
27 de octubre, 2023

Modelo de Machine Learning para la eficiencia operativa del portafolio de productos y servicios del área de Cultura, Recreación y deporte de una Caja de Compensación en Colombia

**GIOVANNY ROBERTO GONZÁLEZ PEDRAZA
JUAN CAMILO CIFUENTES RUIZ**

Trabajo de grado presentado como requisito para optar al título de:

Magister en Inteligencia de Negocios

Directora:

HAILY JOHANNA MORENO CEBALLOS

Modalidad:

Trabajo Dirigido

“Business case”

Universidad EAN

Facultad de Ingeniería

Maestría en Inteligencia de Negocios

Bogotá, Colombia

27 de octubre, 2023

Nota de aceptación:

Firma del jurado

Firma del jurado

Firma del director del trabajo de grado

Bogotá, 07/11/2023

Resumen

En el presente documento, se presenta el diseño de un modelo de machine learning de optimización de tarifas del portafolio de productos y servicios, para mejorar la eficiencia operativa del área de Cultura, Recreación y Deporte de una Caja de Compensación familiar en Colombia. Se incluye la exposición de los objetivos de la iniciativa, su justificación, viabilidad (mediante un análisis de riesgos y de mercado), así como un marco teórico que respalda la estructura del modelo propuesto. Además, se describen los algoritmos utilizados en la solución (Modelo de regresión lineal múltiple, Random Forest, Gradient Vector Forest y Gradient Boosting).

Se aborda el proceso de creación del modelo y los resultados obtenidos, incluyendo un análisis del costo asociado a la implementación de esta iniciativa en la organización. Así mismo, se comparte la validación realizada en colaboración con el departamento responsable, en la que se incluye la exhibición de los productos entregados a la organización (notebook jupyter y dashboard) y se destaca la retroalimentación recopilada de las partes interesadas mediante nuestro instrumento de validación que contuvo análisis tanto cuantitativos como cualitativos. También se describe un plan de intervención para la implementación del modelo propuesto. Por último, se presentan las recomendaciones dadas al área, así como las conclusiones del trabajo realizado.

Palabras clave: Aprendizaje Automático, inteligencia artificial, optimización, modelo de precios, eficiencia operativa

Abstract

This document presents the design of a Machine Learning model for product and service portfolio's price optimization to improve the operational efficiency of the Recreation and Sports area of a Caja de Compensación in Colombia. The document includes the goals, justification, feasibility (through a risk and market analysis), as well as a theoretical framework that supports the structure of the proposed model. In addition, algorithms used in the solution (Multiple Linear Regression Model, Random Forest, Gradient Vector Forest and Gradient Boosting) are described.

The process of creating the model and the results obtained are discussed, including an analysis of the costs associated with the implementation. This work also shares the validation performed in collaboration with the responsible department, including the exhibition of the Minimum Viable Product (notebook jupyter and dashboard) and highlights the feedback collected from stakeholders through our validation form, including both quantitative and qualitative analysis. A balanced score card for the implementation plan is also illustrated. Finally, present the recommendations as well as the conclusions of the business case.

Keywords: Machine Learning, Artificial Intelligence, optimization, pricing, operational efficiency

Contenido

1. Objetivos y alineación estratégica	10
1.1 Objetivo general.....	10
1.2 Objetivos específicos.....	10
2. Justificación y antecedentes del proyecto	11
3. Viabilidad y beneficios esperados.....	13
3.1 Beneficios Esperados	13
3.2 Análisis Financiero.....	14
3.3 Riesgos	16
4. Marco de referencia.....	18
4.1 Usos del Machine Learning.....	18
4.2 Tipos de Machine Learning.....	18
4.3 Analítica de datos	19
4.4 Fases del proceso de Machine Learning.....	20
4.5 Modelo de Madurez Operacional	21
4.6 Modelo de Pricing	22
4.6.1 Regresión Lineal.....	22
4.6.2 Random Forest	22
4.6.3 Gradient Boosting.....	22
4.6.4 Support Vector Regression.....	23

5. Evaluación del mercado	24
5.1 Análisis Externo	24
5.2 Análisis Interno.....	27
5.2.1 Población muestra y ficha técnica.....	27
5.2.2 Identificación de las variables	28
5.2.3 Instrumento de Medición.....	28
6. Plan de Intervención	30
6.1 Cuadro de Mando Integral	30
6.2 Ejemplo de Implementación	30
6.3 Resultado Modelo.....	32
Conclusiones y Recomendaciones.....	33
Conclusiones	33
Recomendaciones.....	33
Referencias	35

Lista de Figuras

Ilustración 1. Mapa de riesgo de impacto y probabilidad. Elaboración propia.....	16
Ilustración 2. <i>Fases de un Proceso de Machine Learning.. Tomado de Las 7 Fases del Proceso de Machine Learning, por (Heras, 2020)</i>	21
Ilustración 3. Análisis Cualitativo y Cuantitativo Cuestionario Validación. Elaboración propia en PowerBI.....	29
Ilustración 4. BSC perspectiva finanzas. Elaboración propia en Miro	30
Ilustración 5. Notebook Resultado predicción Modelo de Machine Learning-. Elaboración propia en Jupyter	31
Ilustración 6. Reporte Gerencial Resultados Modelo de Machine Learning-. Elaboración propia en PowerBI.....	32

Lista de Tablas

Tabla 1. KPI Valor Actual Neto.....	13
Tabla 2. KPI Tasa Interna de Retorno.....	13
Tabla 3. KPI payback.....	14
Tabla 4. Costos proyecto.....	15
Tabla 5. Flujo de caja libre proyecto.....	15
Tabla 6. Matriz DOFA.....	24
Tabla 7. Estrategias Diagnóstico DOFA.....	26
Tabla 8. Ficha Técnica cuestionario.....	27
Tabla 9. Tipo de variables del modelo.....	28

1. Objetivos y alineación estratégica

1.1 Objetivo general

Diseñar un modelo de Machine Learning para la eficiencia operativa de productos y servicios del negocio de cultura, recreación y deporte de una Caja de Compensación en Colombia

1.2 Objetivos específicos

OE1: Explorar mejores prácticas para la implementación de modelos de Machine Learning

OE2: Diagnosticar los factores clave de éxito para la implementación de un modelo de Machine Learning en el sector de bienestar en el área de cultura, recreación y deportes

OE3: Validar con las partes interesadas la implementación de un modelo de Machine Learning en una Caja de Compensación en Colombia

OE4: Establecer un plan de acción para la implementación del modelo Machine Learning propuesto.

2. Justificación y antecedentes del proyecto

La aplicación de conocimientos avanzados en áreas clave de la analítica de datos se ha convertido en un diferenciador estratégico. Estos conocimientos, adquiridos y consolidados a través de la maestría, abarcan aspectos de analítica organizacional, modelos estadísticos, estrategias de Big Data y técnicas para la inteligencia de negocios. La creciente importancia del aprendizaje automático o Machine Learning en este contexto es innegable. Este enfoque permite transformar datos en información estratégica, mejorando la toma de decisiones al identificar patrones y tendencias ocultas en datos complejos.

Además, el Machine Learning facilita la eficiencia operativa al generar planes estratégicos basados en mediciones oportunas y análisis de datos precisos. Sin embargo, para que estas capacidades de Machine Learning sean efectivas, es esencial comprender el entorno en el que operan las organizaciones. Esto es especialmente crítico en sectores altamente dinámicos y competitivos, como el de las cajas de compensación. Aquí es donde la aplicación de modelos de Machine Learning cobra un papel estratégico. Estos modelos no solo ayudan a comprender y adaptarse a este entorno, sino que también ofrecen la capacidad de tomar decisiones informadas y eficaces, lo que se traduce en una ventaja competitiva.

Las cajas de compensación se ven constantemente influenciadas por factores económicos, sociales, tecnológicos y regulatorios. La innovación y la adaptación son imperativas para mantenerse competitivas y relevantes en este entorno cambiante. En el informe de la Supersubsidio (Informe SuperSubsidio,2022). donde se aplicó una encuesta a 43 cajas, se obtuvo que el 33% de las cajas no tienen aplicación en Inteligencia Artificial, el 53% tienen de 1 a 3 opciones y solo el 14% tienen más de 4 aplicaciones. El informe de la Supersubsidio revela un panorama donde la implementación de inteligencia artificial en el sector aún no se ha explorado en su totalidad, lo que representa una oportunidad valiosa para una Caja de Compensación en Colombia. Los riesgos y desafíos, tanto internos como externos, exigen una visión clara y la adopción de estrategias innovadoras. El informe muestra la viabilidad, factibilidad y agregación de valor que tendría una compañía o sector al definir bien un modelo de analítica de datos con enfoque Machine Learning para tomar decisiones.

En el caso específico de una Caja de Compensación en Colombia, una entidad sólida en el mercado, la necesidad de avanzar hacia una gestión basada en datos y análisis se ha vuelto evidente. A pesar de contar con diversos canales de atención al cliente, aún no se ha aprovechado plenamente

la vasta cantidad de datos y sistemas de información para la toma de decisiones estratégicas. Por lo tanto, este proyecto de implementación de un "Modelo de Machine Learning para la eficiencia operativa del portafolio de productos y servicios del área de Cultura, Recreación y Deporte de una Caja de Compensación en Colombia" surge como una iniciativa estratégica.

Se busca aplicar los conocimientos en analítica de datos y Machine Learning para optimizar específicamente el área de servicios de cultura, recreación y deporte de la organización. Este enfoque permitirá a una Caja de Compensación en Colombia explorar prácticas óptimas, diagnosticar factores clave de éxito y validar la implementación de un modelo de Machine Learning en este sector específico. Además, se establece un plan de acción que orientará la implementación efectiva de este modelo para mejorar la eficiencia operativa y la calidad de los servicios ofrecidos en el área de cultura, recreación y deporte de una Caja de Compensación en Colombia.

3. Viabilidad y beneficios esperados

Con el presente proyecto se busca ofrecer a una Caja de Compensación en Colombia una serie de beneficios, incluyendo: Optimización de ingresos, Mejora de la eficiencia operativa, la productividad y la satisfacción del cliente.

3.1 Beneficios Esperados

Las ventajas del proyecto sobre alternativas opcionales como el modelo tradicional de tarifas costo-beneficio son: la toma de decisiones basada en datos, automatización del proceso reduciendo errores humanos, reducción de tareas repetitivas, más agilidad y flexibilidad de cambios, reducción de costos en actividades como estudio de precios- mercadeo o reprocesos. A continuación, se describen los objetivos financieros del proyecto

Tabla 1. KPI Valor Actual Neto

VAN	
Objetivo	Alcanzar un Valor Actual Neto del área de cultura, recreación y deportes en los próximos 3 años
KPI	Beneficio Neto Actualizado (BNA) – Inversión Inicial (I ₀)
Meta	>=\$100 millones
Tiempo	Anual
Propietario	Director Financiero

Nota: Elaboración propia

Tabla 2. KPI Tasa Interna de Retorno

TIR	
Objetivo	Aumentar la Tasa Interna de retorno del área de cultura, recreación y deportes en los próximos 5 años
KPI	$[(\text{Valor final} - \text{valor inicial}) / \text{Valor inicial}] \times 100$
Meta	>1.5%

Tiempo	Anual
Propietario	Director Financiero

Nota: Elaboración propia

Tabla 3. KPI payback

Payback	
Objetivo	Recuperar la inversión en el proyecto en los próximos 5 años
KPI	Inversión inicial / Flujo de efectivo anual
Meta	100% dinero recuperado
Tiempo	Mensual
Propietario	Director Financiero

Nota: Elaboración propia

3.2 Análisis Financiero

Se realiza la identificación de los ingresos, gastos y costos de la puesta en marcha del proyecto propuesto, así como las proyecciones y los incrementos para cinco años, por otro lado, los recursos necesarios para su ejecución como lo son: equipos, infraestructura, personal capacitado. Se toma como base los salarios promedios en Colombia para los profesionales requeridos según reportes del sector como por ejemplo (Talent - Salario Medio, 2023) o también (Indeed,2023). Así mismo, para el costo de adquisición (TCO), en los reportes de los principales proveedores de cloud como Amazon (AWS SageMaker TCO,2020) o el informe de expertos (Coop, 2021).

Los supuestos para realizar la proyección del proyecto, así como su rentabilidad se dan en base a los datos proyectados por el área de investigaciones económicas de Bancolombia, para el caso del IPC se asumen los valores del informe (Bancolombia - Proyecciones económicas Colombia, 2023). La Tasa de impuesto de renta es la estipulada para el año 2023 de un 35%. Por otro lado, la tasa del crédito que se asume es la otorgada por Bancolombia en su página para empresas de un 23,94% E.A. Y que la tasa de inversión sea de un 13% dado el crecimiento que se espera de la economía para los próximos 4 años.

Dada la naturaleza tecnológica del proyecto se establecen los siguientes servicios que van a generar costos en la creación y ejecución del modelo durante el primer año.

Tabla 4. Costos proyecto

Nombre del servicio	Cantidades	Costo unitario del servicio	Costos totales
Ingeniero de datos	1	\$ 5.500.000,00	\$ 5.500.000
Gerente de proyectos	1	\$ 10.000.000,00	\$ 10.000.000
Científico de datos	2	\$ 5.000.000,00	\$ 10.000.000
Activos tecnológicos	3	\$ 3.000.000,00	\$ 9.000.000
Infraestructura tecnológica	1	\$ 20.000.000,00	\$ 20.000.000
Beneficios Positivos Modelo	1	\$ 7.000.000,00	\$ 7.000.000
		TOTAL	\$ 61.500.000

Nota: Elaboración propia

Como el proyecto es una intervención empresarial, la empresa asume los costos fijos de oficina y por tanto no se requiere inversión en arriendo o compra de inmuebles, ni gastos de papelería o mercadeo. Así mismo, no se contemplan equipos de transporte ni adquisición de franquicias o licencias de software. El costo de adquisición son el pago por uso de los modelos en la infraestructura del proveedor cloud y el costo de nómina del recurso humano empleado.

El VPN del proyecto es de \$24'580.311 millones de pesos con una Tasa interna de retorno del 20,63%, durante un periodo de recuperación esperado de 4.12 años. La tabla 5 ilustra el flujo de caja libre del proyecto

Tabla 5. Flujo de caja libre proyecto

CALCULO DEL FLUJO DE CAJA LIBRE					
EBIT	\$	\$	\$	\$	\$
	27.200.000,0	32.630.000,0	34.860.515,0	35.917.362,4	36.761.206,2
Impuestos	\$	\$	\$	\$	\$
	9.520.000,0	11.420.500,0	12.201.180,3	12.571.076,8	12.866.422,2

NOPLAT	\$	\$	\$	\$	\$
	17.680.000,0	21.209.500,0	22.659.334,8	23.346.285,5	23.894.784,0
Inversión Neta	\$	\$	\$	\$	\$
	14.818.138,6	11.884.041,3	17.849.217,5	23.232.294,6	29.097.014,2
Flujo de Caja	\$	\$	\$	\$	\$
Libre del periodo	32.498.139	33.093.541	40.508.552	46.578.580	52.991.798

Nota: Elaboración propia

El detalle del análisis financiero se encuentra en el *Anexo1 SIMULADOR FINANCIERO SIMPLIFICADO.xls*

3.3 Riesgos

Se realiza un análisis de los riesgos que se pueden presentar en la ejecución del proyecto, se procede con una evaluación cualitativa y cuantitativa que permita evaluar el impacto y la probabilidad de ocurrencia para la toma de acciones. La figura 1 muestra la matriz de clasificación de los riesgos.

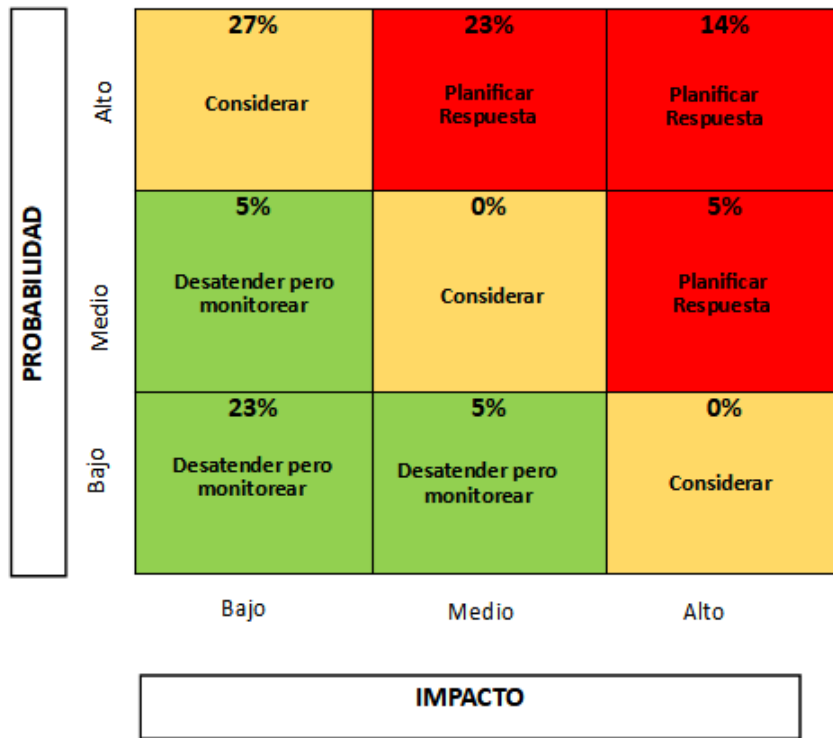


Ilustración 1. Mapa de riesgo de impacto y probabilidad. Elaboración propia

De los riesgos identificados y evaluados para el proyecto se destacan que alrededor del 40% son de nivel Alto y requieren un monitoreo que permita planificar una respuesta a tiempo. Un 30% son de nivel Medio que deben ser considerados para monitoreo, y los restantes son de nivel Bajo y pueden ser desatendidos inicialmente. En el *Anexo 2. Análisis Riesgos.xlsx* se encuentra el detalle de los riesgos y su evaluación de probabilidad, impacto de presencia y acciones con su plan de respuesta.

4. Marco de referencia

Este documento identifica en la literatura, modelos, metodologías y marcos de trabajo que ayuden a identificar y analizar aquellas variables a tener en cuenta en este Business Case para el desarrollo y evaluación del modelo de Machine Learning para la eficiencia operativa del portafolio de productos y servicios de Una Caja de Compensación en Colombia en su iniciativa de transformación y adopción de Inteligencia Artificial. El aprendizaje automático o Machine Learning (ML) en palabras de (Joyanes, 2019) busca entrenar algoritmos que se alimentan de históricos para construir modelos estadísticos y computacionales que permitan a las computadoras simular un aprendizaje basado en datos que habiliten la generación de predicciones.

4.1 Usos del Machine Learning

El Machine Learning (ML) se ha convertido en una herramienta clave en diversas áreas, desde la investigación científica hasta la industria y el comercio. Algunos de los principales usos son:

Reconocimiento de patrones y clasificación: Aplicados a la clasificación y el reconocimiento de patrones en datos, como modelos de regresión y clasificación lineal, redes neuronales, métodos y máquinas de Kernel (Bishop, 2006)

Procesamiento de imágenes y visión por computadora: (Krizhevsky et al., 2012) Presentan un enfoque basado en redes neuronales convolucionales para la clasificación de imágenes, que demostró un rendimiento superior en la tarea de clasificación de imágenes en comparación con otros enfoques.

Análisis de datos y predicción: El machine learning también se utiliza para el análisis de grandes conjuntos de datos y la predicción de resultados en diversas áreas, como las finanzas y la medicina. (Witten et al., 2016)

Procesamiento del lenguaje natural: Es un área de la inteligencia artificial que se enfoca en la comprensión del lenguaje humano por parte de las computadoras. Se utiliza ampliamente en esta área para el análisis semántico y la clasificación de texto. (Jurafsky & Martin, 2020)

4.2 Tipos de Machine Learning

Dentro del ML existen algoritmos de aprendizaje supervisado (Joyanes, 2019) que requieren la intervención humana para identificar y etiquetar los datos de entrenamiento, aprendizaje no supervisado que no requieren del etiquetado de los datos y aprendizaje reforzado que es un híbrido entre el supervisado y el no supervisado donde se realiza una retroalimentación basada en recompensa.

Aprendizaje Supervisado: Se utilizan conjuntos de datos etiquetados para entrenar algoritmos que clasifiquen datos o predigan resultados de forma precisa (Verdhan, 2020). Utiliza un conjunto de datos de entrenamiento para enseñar a los modelos a generar la salida deseada. Este conjunto de datos incluye datos de entrada y resultados correctos, que permiten que el modelo aprenda con el tiempo. Algunos de los algoritmos más comunes en el aprendizaje supervisado son: Árboles de decisión, Clasificación de Naïve Bayes, Regresión por mínimos cuadrados, Regresión Logística, Support Vector Machines (SVM), Métodos “Ensemble” (Conjuntos de clasificadores)

Aprendizaje No Supervisado: Se distingue del aprendizaje supervisado por el hecho de que no hay un conocimiento a priori, analiza y agrupa conjuntos de datos no etiquetados. Estos algoritmos descubren patrones ocultos o agrupaciones de datos sin necesidad de intervención humana. (Kyan, 2014) Se utilizan para tres tareas principales: agrupamiento, asociación y reducción de dimensionalidad Algunos ejemplos de algoritmos son: K-means, Agrupamiento jerárquico, Reducción de dimensionalidad

Aprendizaje Reforzado: Es un área inspirada en la psicología conductista, cuyo fin es determinar qué acciones debe escoger un agente de software en un entorno dado con el fin de maximizar alguna noción de "recompensa" o premio acumulado. (Sutton & Barto, 2018). Un agente interactúa con su entorno y recibe recompensas o penalizaciones por sus acciones. El objetivo del agente es aprender a tomar decisiones que maximicen la recompensa acumulada a lo largo del tiempo. Esta retroalimentación le permite al agente evaluar la calidad de sus acciones y ajustar su comportamiento en consecuencia.

4.3 Analítica de datos

La analítica de datos es el proceso de examinar conjuntos de datos para sacar conclusiones sobre la información que contienen, con el objetivo de mejorar la toma de decisiones. (Evans, 2020) clasifica los tipos de analítica en:

Análisis descriptivo: este tipo de análisis se utiliza para resumir y describir los datos. El objetivo es proporcionar una comprensión general de los datos y su distribución. El análisis descriptivo incluye técnicas como la visualización de datos, los resúmenes estadísticos y las tablas de frecuencia.

Análisis predictivo: el análisis predictivo se utiliza para predecir el comportamiento futuro de un sistema en función de los datos históricos. El objetivo es encontrar patrones y tendencias en los datos que puedan utilizarse para hacer predicciones. Puede predecir riesgos y encontrar relaciones en los datos que no son fácilmente perceptibles con análisis tradicionales.

Análisis prescriptivo: el análisis prescriptivo se utiliza para proporcionar recomendaciones y soluciones específicas a problemas de negocio. Utiliza técnicas como la optimización y la simulación para identificar la mejor alternativa para maximizar o minimizar un objetivo. Es usada en muchas áreas de negocio como finanzas y mercadeo, por ejemplo, ayuda a determinar el mejor precio para maximizar las ganancias, o seleccionar el mejor portafolio de inversión

4.4 Fases del proceso de Machine Learning

Basado en el modelo CRISP-DM (Cross-Industry Standard Process for Data Mining) que es una metodología (Hotz, 2023) flexible que incluye un ciclo vital por fases, (Heras, 2020) propone un modelo de 7 fases para el desarrollo del proceso de Machine Learning:

Entender el Problema: Es importante entender el problema que se quiere resolver y colaborar con las personas que saben mucho acerca del problema

Definir un Criterio de Evaluación: Se define cómo se evaluará el modelo desarrollado,

Evaluación de la solución actual: Se evalúa la solución actual al problema.

Preparar los datos: Cubre todas las actividades para construir el conjunto de datos definitivo que se empleará en la siguiente fase

Construir el modelo: En esta fase se desarrolla la solución al problema de negocio planteado.

Análisis de Errores: En esta fase se analizan los errores del modelo desarrollado.

Modelo integrado en un Sistema: En esta fase se implementa el modelo desarrollado en un sistema

La figura 2 ilustra el proceso y sus fases

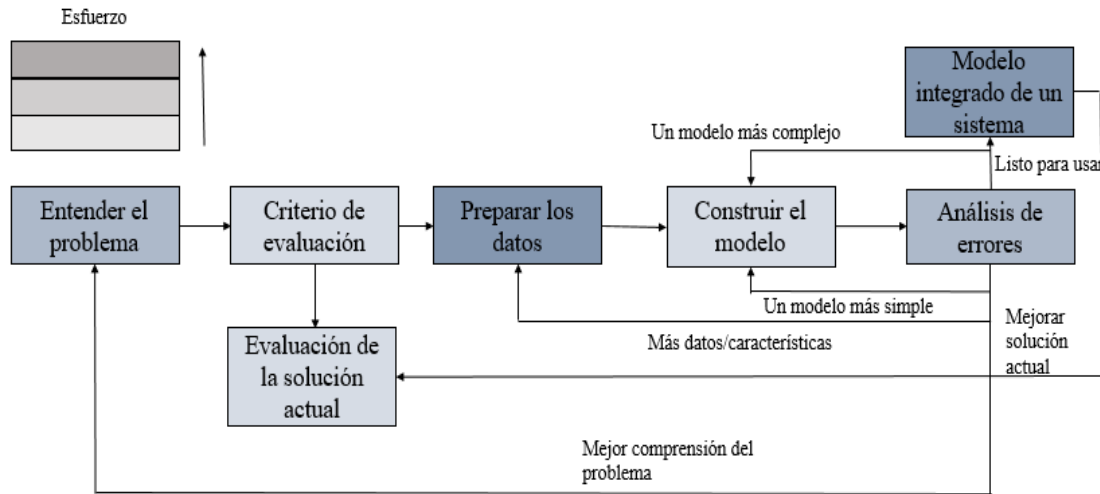


Ilustración 2. Fases de un Proceso de Machine Learning. Tomado de *Las 7 Fases del Proceso de Machine Learning*, por (Heras, 2020)

4.5 Modelo de Madurez Operacional

El modelo de madurez de operaciones de Machine Learning (Microsoft- Azure Architecture Center, s.f.) ayuda a establecer prácticas y los principios de las operaciones de desarrollo (DevOps) necesarios para ejecutar un entorno correcto de Machine Learning (MLOps). Muestra la mejora continua en la creación y el funcionamiento de un entorno de aplicación de ML a nivel de producción. Muestra cómo aumentar la capacidad de MLOps de manera incremental. El modelo MLOps define cinco niveles:

Nivel 0 - Sin MLOps: Ciclo de vida del modelo de Machine Learning completo difícil de administrar.

Nivel 1 - DevOps sin MLOps: Los modelos se implementan en producción sin pruebas ni validación.

Nivel 2 - MLOps básico: Los modelos se prueban y validan antes de la implementación en producción.

Nivel 3 - MLOps avanzado: Los modelos se prueban y validan automáticamente antes de la implementación en producción.

Nivel 4 - MLOps maduro: Los modelos se prueban y validan automáticamente antes de la implementación en producción con pruebas avanzadas como A/B testing y shadow testing.

4.6 Modelo de Pricing

Un modelo de pricing (Talluri, y Van Ryzin, 2004) es un modelo basado en Machine Learning para recomendar precios para productos o servicios. Se entrena a partir de datos históricos, como precios, ventas y demanda, para predecir el precio óptimo para un determinado producto o servicio. Suelen utilizar algoritmos de optimización para encontrar el precio óptimo. Su principal es su capacidad para procesar volúmenes altos de datos y predecir la demanda futura con mayor precisión (Gallego y Conejo, 2015). Algunos de los algoritmos o técnicas utilizados para la implementación de estos modelos son:

4.6.1 Regresión Lineal

La regresión lineal es una técnica estadística utilizada para modelar la relación entre una variable dependiente (o respuesta) y una o más variables independientes (o predictoras) mediante una función lineal. (Gujarati y Porter, 2022). Para analizar la regresión lineal, se puede utilizar los siguientes estadísticos (Wooldridge, 2021): Coeficiente de determinación (R^2), Error estándar de estimación (RMSE), P-valor. La regresión lineal se utiliza en una amplia variedad de campos, incluyendo la economía, la psicología, la ciencia de datos, la ingeniería y muchas otras disciplinas.

4.6.2 Random Forest

Un bosque aleatorio (IBM, s.f) es una técnica de Machine Learning empleada para la clasificación y la regresión. Es una técnica popular entre los científicos de datos por su gestión de grandes volúmenes de datos (Breiman, 2001). Su mecanismo es utilizar varios árboles de decisión. Cada árbol decide según las propiedades del conjunto de datos y devuelve un resultado, luego, el bosque aleatorio combina los resultados de todos los árboles para producir una predicción final. Para evaluar el rendimiento de un bosque aleatorio (Kreiger, 2020), se pueden utilizar varias métricas, como, por ejemplo: Precisión, Sensibilidad, Especificidad, Valor F, Curva ROC.

4.6.3 Gradient Boosting

El reforzamiento de gradiente (Friedman, 2001) se basa en la construcción de un conjunto de modelos de aprendizaje automático simples, que se combinan para producir un modelo más complejo y preciso.

El gradient boosting funciona de la siguiente manera (Sklearn Gradient Boosting Classifier, s/f): primero se crea un modelo de Machine Learning sencillo. Segundo se realiza el cálculo del error del modelo en los datos de entrenamiento. Tercero se crea un modelo nuevo orientado a subsanarlos

errores del anterior modelo. Finalmente se repiten los pasos 2 y 3 hasta que se alcanza el objetivo deseado.

4.6.4 Support Vector Regression

El algoritmo SVR es utilizado para el análisis de regresión, busca encontrar una función no lineal que pueda predecir valores numéricos continuos a partir de datos de entrada. SVR es una extensión de Support Vector Machines (SVM) que se utiliza principalmente en problemas de clasificación (Vapnik, 1995). Utiliza vectores de soporte, un hiperplano de regresión con un margen ϵ y funciones de núcleo para encontrar la mejor función de regresión en un espacio de mayor dimensión. SVR es útil cuando se trata de problemas de regresión no lineales y cuando se necesita un enfoque que tenga en cuenta los valores atípicos en los datos.

5. Evaluación del mercado

Este trabajo es una investigación aplicada, ya que se obtendrán entregables que cumplan con los objetivos planteados. Así mismo, el grado de profundidad es descriptivo de acuerdo con el nivel de información técnico de esta intervención empresarial. Adicionalmente, es una investigación cuantitativa ya que es necesaria la recolección de información teórica, hacer una evaluación y diagnóstico de la organización y desarrollar un modelo basado en datos. En este orden de ideas, esta es una investigación cuantitativa. Por otro lado, es una investigación deductiva, pues se tomarán los componentes principales del marco teórico para aplicarlos en un caso particular de Una Caja de Compensación en Colombia. Finalmente, la investigación es transversal, teniendo en cuenta los requerimientos de la organización, se ejecutará de manera continua y no por intervalos.

5.1 Análisis Externo

El análisis externo de una Caja de Compensación en Colombia se realizó con el Modelo DOFA que permite identificar cómo el entorno impacta a la organización. Se analizarán los diferentes elementos del modelo **DOFA** y de qué manera afectan a una Caja de Compensación en Colombia. La tabla 6 muestra el modelo DOFA con la descripción de los aspectos más relevantes por factor relacionados con la mejora operativa del área de cultura, recreación y deportes de una Caja de Compensación en Colombia.

Tabla 6. Matriz *DOFA*

MATRIZ DOFA				
No.	ANALISIS INTERNO		ANALISIS EXTERNO	
	FORTALEZAS	DEBILIDADES	OPORTUNIDADES	AMENAZAS
1	El conocimiento del negocio.	Falta de experiencia en la utilización de Machine	Optimización en los precios de las tarifas ofertadas a los clientes.	Obsolescencia modelo

		Learning para la toma de decisiones.		
2	Experiencia y reputación.	Bajo nivel de madurez de analítica.	Precios objetivos dado las características sociales, tributarias y financieras.	Cambios en la normativa de ley en las cajas de compensación
3	Datos de corte transversal con variables que determinan la Tarifa.	Tomas de decisiones financieras con modelos clásicos	Creación de nuevos productos y servicios.	Competidores
4	Datos estructurados.	Presupuesto limitado	Propuestas de más modelos para la toma de decisiones de tarifas de los productos.	Disminución oferta de usuarios
5	Segmentación optima de productos.	Infraestructura no adecuada	Mejoramiento de la eficiencia operativa en la toma de decisiones.	Cambios en tecnología
6	Información heterogénea de grupos de clientes, tarifas, costos, impuestos.			

Nota: Elaboración propia

Teniendo en cuenta la calificación dada a cada elemento de la matriz DOFA y la identificación de las estrategias (ver *Anexo3 Análisis DOFA.xlsx*), se seleccionan las mejores estrategias que podrían mejorar las condiciones actuales de una Caja de Compensación en Colombia permitiendo

mejorar su participación en el mercado, y aprovechando su fortaleza y competitividad. La tabla 7 describe las estrategias seleccionadas

Tabla 7. Estrategias Diagnóstico DOFA

	ESTRATEGIAS FO	EXPLICACION
1	F1O1	El conocimiento del negocio nos permite optimizar los precios de las tarifas de bienes y servicios
2	F2O2 F5O4	Aumento de oferta
3	F3O4	Introducción de nuevos productos a nuevos mercados
4	F4O4	Oportunidad de mejora en la toma de decisiones operativas del área
	ESTRATEGIAS FA	
1	F1A1	Ampliación y mejoramiento de la oferta de productos y servicios
2	F2A2A3	Generar asociación positiva y generar valor de marca
3	F5A3	Exportar y desarrollar nuevos mercados
4	F6A5	Exportar y desarrollar nuevos mercados y productos
5	F1A2A3	Buscar mayores ventas mejorando(desarrollando) el producto actual. o modificándolo
	ESTRATEGIAS DO	
1	D1O1	Optimización de tarifas en los productos y servicios.
2	D1O1O2	Desarrollo de estrategias de la gestión comercial y reestructuración de las ventas
3	D4O3O4	Tener tecnología que soporte la infraestructura de modelos ML mejorara la optimización de estos haciendo más eficiente las decisiones
	ESTRATEGIAS DA	
1	D1A1	Utilización de Machine Learning para la optimización operativa

2	D3A3	Tener un presupuesto adecuado para la incorporación de modelos a producción aumentara las ventajas competitivas de la empresa
3	D4A1A4	Tener escalabilidad tecnológica para tener sostenibilidad de la innovación

Nota: Elaboración propia

5.2 Análisis Interno

En esta sección se presenta el análisis interno de la organización para el modelo propuesto

5.2.1 Población muestra y ficha técnica

La muestra está conformada por un grupo que corresponde a las personas cuyo rol y actividades diarias están asociadas y tienen influencia en los procesos de la operatividad del área de cultura, recreación y deportes

Tabla 8. Ficha Técnica cuestionario

Característica	Descripción
Ciudad	Medellín, Antioquia
Población	Directores, Gerente de unidad, profesionales TI y negocio.
Muestra	Directivos (3), Gerente de Unidad (1) y Profesionales TI (6)
Confianza	NA
Medio	Formulario Web
Periodo	Octubre 2023
Selección	Segmentada.
Roles	Analista, directivo Gerente de Área

Nota: Elaboración propia

5.2.2 Identificación de las variables

Con base en el marco teórico se logran identificar variables significativas para la estimación de tarifas (pricing) para la gestión en la toma de decisiones basado en la analítica. Las variables identificadas para determinar las tarifas a nivel organizacional son de naturaleza cuantitativa y cualitativa, la data utilizada fue proporcionada por la organización dado los temas de experto:

Tabla 9. *Tipo de variables del modelo*

Variable	Tipo
Tipo de tarifa	Cualitativa ordinal
Código Sublínea	Cualitativa ordinal
Costo	Cuantitativa continua
Intensidad Horaria	Cuantitativa discreta
No mínimo de participantes	Cuantitativa discreta
No máximo de participantes	Cuantitativa discreta

Nota: Elaboración propia

5.2.3 Instrumento de Medición

Para la validación de la propuesta de intervención se utilizó un cuestionario tipo encuesta que permitió identificar la validación del prototipo de Machine Learning para el área de cultura, recreación y deporte de dicha entidad. El instrumento de validación está creado para ciertos grupos de la caja de compensación como lo son: Directivos (4), Gerente unidad de negocio (1) y expertos tecnológicos del área de la unidad de negocio (6). El instrumento está conformado por preguntas abierta y cerradas que nos permitirán un análisis cuantitativo y cualitativo de nuestro prototipo. La primera parte de nuestro instrumento es la introductoria es la cual preguntamos cargo, tiempo en la compañía, experiencia entre otras. Para la segunda parte hacemos unas breves preguntas diagnósticas y de retos que enfrenta el área de negocios. Por último, la tercera parte realiza la validación de nuestra propuesta que nos llevará a validar nuestros objetivos y pasos a seguir por parte de la compañía. Ver *Anexo 4. Cuestionario Validación.xlsx*

La figura 3 muestra un tablero de control con los resultados del cuestionario de validación. (Ver *Anexo 5. Informe Análisis Datos.pbix*)

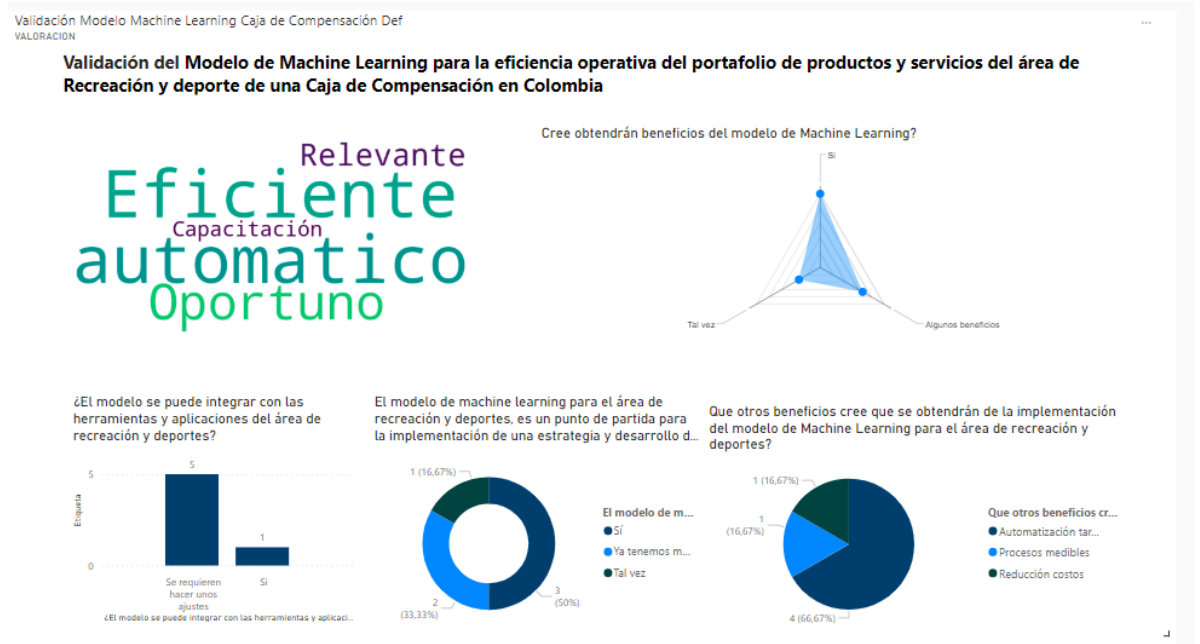


Ilustración 3. Análisis Cualitativo y Cuantitativo Cuestionario Validación. Elaboración propia en PowerBI

Puede acceder al informe completo en el siguiente enlace:

https://app.powerbi.com/links/5rR8G9zscX?ctid=c06fb559-1b68-4b84-a14f-47d0d837a5ab&pbi_source=linkShare

6. Plan de Intervención

Como parte de los objetivos de este business case, se establece un plan de intervención para la implementación del modelo Machine Learning propuesto.

6.1 Cuadro de Mando Integral

Se realiza un cuadro de mando integral (CMI) o Balanced ScoreCard (BSC), que incluye tanto objetivos estratégicos como financieros de la organización a través de diferentes dimensiones donde el foco central son la visión y la estrategia. La figura 4 ilustra el BSC para la perspectiva finanzas. Puede ver el cuadro completo en el *Anexo 6. BSC.pdf*

Finanzas		
Objetivo	Meta	Indicador
Aumentar ingreso del área de cultura, recreación y deportes	=> a el 1% en 5 años	$TIR = (\text{Valor final} - \text{valor inicial}) / \text{Valor inicial} \times 100$
Alcanzar un Valor Actual Neto del área de cultura, recreación y deportes	$VAN \geq 100$ millones en 3 años	$VAN = \text{Beneficio Neto Actualizado (BNA)} - \text{Inversión Inicial (I}_0)$
Recuperar la inversión	≤ 5 años	$\text{Payback} = \text{Inversión inicial} / \text{Flujo de efectivo anual}$

Ilustración 4. BSC perspectiva finanzas. Elaboración propia en Miro

6.2 Ejemplo de Implementación

Siguiendo la metodología CRISP-DM primero se busca entender el negocio y el proceso que realizan en el área de cultura, recreación y deporte de Una caja de compensación en Colombia. En

esta labor se identifica que es posible reducir los tiempos de estimación de las tarifas de los cursos de cultura, recreación y deporte, este proceso tarda aproximadamente entre 2-3 semanas de estimación de dichas tarifas. En la segunda etapa se analiza la estructura de datos y su naturaleza, además de la forma en que los utilizan para dicha estimación de tarifas, en él se evidencia que se utilizan datos estructurados de creación interna. Después de analizar esto junto con el equipo de negocio y TI se define la creación de un modelo de Machine learning de optimización de tarifas que contribuye a mejorar la eficiencia operativa. Dado la data entregada por la caja de compensación y variables discutidas en un comité técnico para la elaboración del modelo se empieza la identificación de las variables para entrenamiento y con ella el tratamiento de las mismas. Se identifican 2 variables cualitativas y 4 cuantitativas. Después de realizar la estandarización y tratamiento de estas variables se escogen 4 modelos la regresión lineal múltiple, Random Forest, Gradient Boosting y Support Vector Regression. Como despliegue se elige el lenguaje de programación Python en el notebook Jupyter. Ver Anexo 7. *Modelo Pricing ML.ipynb*

La figura 5 muestra la salida del modelo con la predicción.

```
Predicciones y evaluación del modelo:
Para cada modelo, se realizan predicciones en el conjunto de prueba.
Se calcula el Error Cuadrático Medio (MSE) y el Coeficiente de Determinación (R^2) para evaluar el rendimiento de cada modelo
en el conjunto de prueba.
Se imprimen los resultados de las métricas de evaluación.

In [11]: for model_name, model in models.items():
        model.fit(X_train, y_train)
        y_pred = model.predict(X_test)
        predictions[model_name] = y_pred

        # Evaluación del modelo
        mse = mean_squared_error(y_test, y_pred)
        r2 = r2_score(y_test, y_pred)
        print(f'Modelo: {model_name}')
        print(f'Error cuadrático medio (MSE): {mse}')
        print(f'Coeficiente de determinación (R^2): {r2}')
        print()

        # Agregar las predicciones al DataFrame original
        for model_name in predictions.keys():
            data[model_name + " Prediction"] = predictions[model_name]

Modelo: Linear Regression
Error cuadrático medio (MSE): 68737047.4719047
Coeficiente de determinación (R^2): 0.9383494435543523

Modelo: Random Forest
Error cuadrático medio (MSE): 1656172.3210717218
Coeficiente de determinación (R^2): 0.9985145718514359

Modelo: Gradient Boosting
Error cuadrático medio (MSE): 1508160.9090192525
Coeficiente de determinación (R^2): 0.9986473239298121

Modelo: Support Vector Regression
Error cuadrático medio (MSE): 2106620473.060643
Coeficiente de determinación (R^2): -0.8894370526616875
```

Ilustración 5. Notebook Resultado predicción Modelo de Machine Learning-. Elaboración propia en Jupyter

En el siguiente enlace se encuentra el entorno de ejecución del modelo propuesto:

<https://colab.research.google.com/drive/1T7WJXjzXtko9IbJnPjCo4stkhLVNO0o?usp=sharing>

Adicionalmente Ver Anexo Modelo Pricing ML, notebook del modelo

6.3 Resultado Modelo

Los resultados sugieren que los modelos Random Forest y Gradient Boosting tienen un rendimiento mucho mejor que el modelo de Regresión Lineal y el modelo SVR. Estos dos modelos tienen valores de R^2 cercanos a 1, lo que indica una buena capacidad para predecir la variable objetivo. El modelo SVR tiene un valor de R^2 negativo, lo que sugiere que no es un buen ajuste para los datos. Ver Anexo 8. *data_con_predicciones.xlsx*

En la figura 6 se muestra el tablero Gerencial con los resultados del modelo

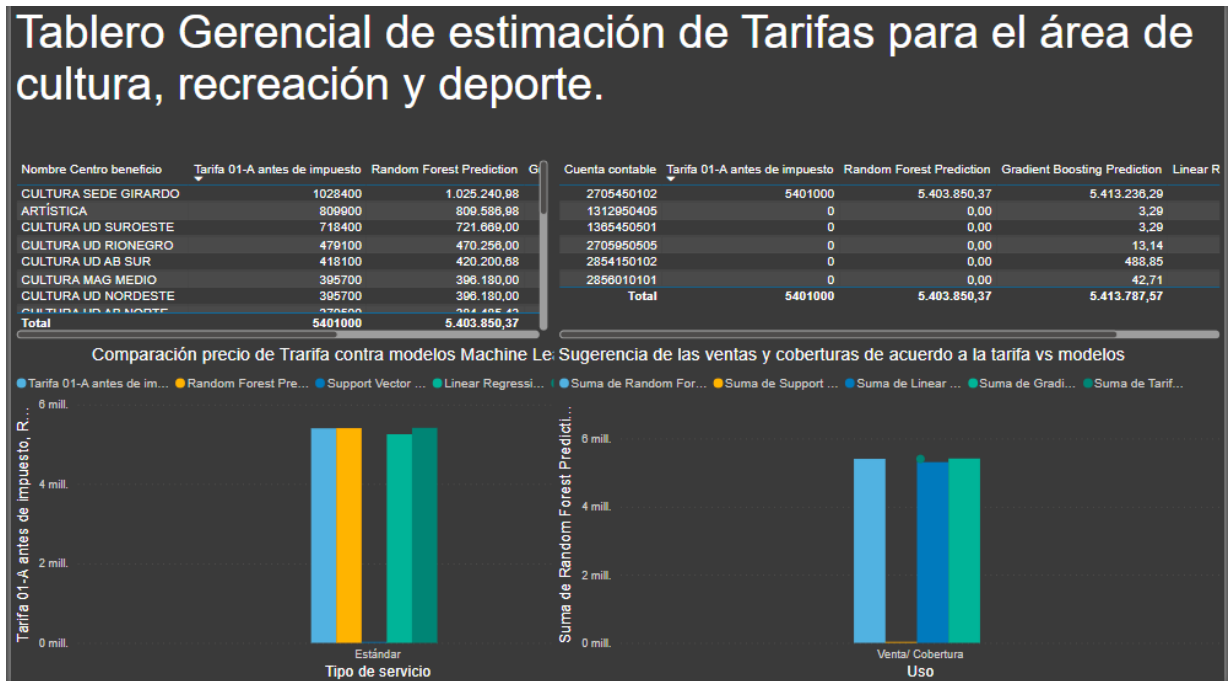


Ilustración 6. Reporte Gerencial Resultados Modelo de Machine Learning-. Elaboración propia en PowerBI

Conclusiones y Recomendaciones

Conclusiones

Se investigaron prácticas y metodologías significativas para la implementación del modelo de Machine Learning como el modelo CRISP-DM y el modelo MLOps que permitieron la consolidación de un modelo para el área de cultura, recreación y deporte de la caja de compensación.

Se identificaron factores clave de éxito como las variables socioeconómicas, económicas y de tarifas para el entrenamiento del modelo de optimización de precios para el portafolio de servicios del área de cultura, recreación y deportes con una respuesta de predicción muy buena en 3 de los cuatro modelos utilizados lo que evidencia el buen comportamiento de los modelos para aprender de la tarea a realizar.

Por medio del cuestionario utilizado como instrumento de medición, se logra validar con las partes interesadas la pertinencia de implementación del modelo propuesto de Machine Learning, que arrojo pilares de mejoramiento a futuro por parte del área impactada y las necesidades que se manejan a nivel de conocimiento de estas tendencias y que apalancaran a futuro el negocio.

Se estableció un plan de intervención para la implementación del modelo de Machine Learning propuesto para el área de cultura, recreación y deportes que nos permitió focalizar por medio de los resultados obtenidos del prototipo e instrumentó de validación los retos a futuro y los beneficios de utilizar Machine Learning para apalancar los objetivos del área.

Se diseñó un modelo de Machine Learning aplicando un modelo de optimización de precios con diferentes modelos de analítica para la eficiencia operativa de productos y servicios del área de cultura, recreación y deporte de una Caja de Compensación en Colombia en el que se evidencia la reducción de costos y tiempos utilizados para dicha tarea.

Recomendaciones

Se recomienda como primera medida empezar a crear data de fuentes externas que sirvan como insumo para focalizar los grupos de interés, así como las tarifas a el problema planteado. Por otro

lado, viendo que se utilizó la transformación de una data, y se finalizó con una visualización se sugiere a el área complementar esta iniciativa con la entrada automática del insumo por medio de sus bases de datos para dar soluciones en tiempo real.

Dado los análisis presentados en la encuesta de validación a la unidad de negocio de la organización se recomienda una “Toma analítica” que busque reforzar los conceptos de analítica básica hasta avanzada en el equipo para determinar si es necesario la adquisición de científicos de datos o esta alguien en la capacidad de adquirir dichos conocimientos para la sostenibilidad de este modelo, así como los futuros modelos que salgan.

Referencias

- Alhashmi, S. F., Alshurideh, M., Al Kurdi, B., & Salloum, S. A. (2020). A systematic review of the factors affecting the artificial intelligence implementation in the health care sector. In Proceedings
- Al-Anqoudi, Y., Al-Hamdani, A., Al-Badawi, M., & Hedjam, R. (2021). Using Machine Learning in Business Process Re-Engineering. *big data and cognitive computing* , 1-19.
- Ali, A., Razak, S. A., Othman, S. H., Al-Dhaqm, A., Nasser, M., Elhassan, T., . . . Saif, A. (2022). Financial Fraud Detection Based on Machine Learning: A Systematic Literature Review. *applied sciences* , 1-24.
- Alpaydin, E. (2020). Introduction to Machine Learning. MIT Press
- AWS Amazon SageMaker. Total Cost of Ownership (TCO) (2020). Awscloud.com. https://pages.awscloud.com/rs/112-TZM-766/images/Amazon_SageMaker_TCO_uf.pdf
- Bancolombia - Actualización de proyecciones económicas Colombia. (2023). Grupobancolombia.com <https://acortar.link/oV9mPs>
- Bianchi, D., De Marco, G., & Kahyaoglu, H. (2019). Deep Learning for Asset Management. *Journal of Risk and Financial Management*, 12(2), 53. doi: 10.3390/jrfm12020053.
- Bishop, C. M. (2006). Pattern recognition and machine learning. Springer
- Breiman, L. (2001). Random forests. Berkeley.edu. <https://www.stat.berkeley.edu/~breiman/randomforest2001.pdf>
- Clemente, A. D. (2020). Modeling Portfolio Credit Risk Taking into Account the Default Correlations Using a Copula Approach: Implementation to an Italian Loan Portfolio. Department of Social and Economic Sciences, “Sapienza” University of Rome, 12-23.
- Cioffi, R., Travaglioni, M., Piscitelli, G., Petrillo, A., & Felice, F. D. (2020). Artificial Intelligence and Machine Learning Applications in Smart Production: Progress, Trends, and Directions. *sustainability*, 1-26.

- Coop, R. (2021). What is the cost to deploy and maintain a machine learning model? phData. <https://www.phdata.io/blog/what-is-the-cost-to-deploy-and-maintain-a-machine-learning-model>
- Di Vaio, A., Palladino, R., Hassan, R., & Escobar, O. (2020). Artificial intelligence and business models in the sustainable development goals perspective: A systematic literature review. *Journal of Business Research*, 121, 283-314.
- Diez-Canedo, J. M. (2006). *Una nueva visión del riesgo de crédito*. Bogotá: LIMUSA NORIEGA EDITORES.
- Dirección de investigaciones Económicas, Sectoriales y de Mercados. (2023). *Resumen - Actualización de proyecciones económicas Colombia*. Grupo Bancolombia.
- Domingos, P. (2018). *The Master Algorithm: How the Quest for the Ultimate Learning Machine Will Remake Our World*. Basic Books.
- EMIS UNIVERSITY (2023). <https://www-emis-com.bdbiblioteca.universidadean.edu.co/php/companies?pc=CO&cmpy=3063657>
- Enholm, I. M., Papagiannidis, E., Mikalefi, P., & Krogstie, J. (2022). Artificial Intelligence and Business Value: a Literature Review. *Department of Computer Science, Norwegian University of Science*, 1709-1734.
- Espinosa, F. R. (2013). Variables financieras determinantes del fracaso empresarial para la pequeña y mediana empresa en Colombia: análisis bajo modelo Logit. *Pensamiento & Gestion*.
- Evans, J. R. (2020). *Business Analytics*. 3er Edition. Pearson.
- Foririsisis Uniandes. (s.f.). Etapas de madurez de una empresa. *INTELIGENCIA DE NEGOCIOS*, 50-52.
- Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *Annals of Statistics*, 29(5), 1189-1232
- Gallego, G., & Conejo, A. J. (2015) *Pricing models in the digital age*. Springer.
- Gan, L., Wang, H., & Yang, Z. (2020). Machine learning solutions to challenges in finance: An application to the pricing of financial products. *Technological Forecasting & Social Change*, 1-11.

- Giraldo, A. L., & Nuñez, M. A. (2021). ADMINISTRACIÓN DEL RIESGO ESTRATÉGICO EN ALGUNAS GRANDES EMPRESAS PRIVADAS DE COLOMBIA . *AD-MINISTER*, 67-96.
- Gujarati, D., & Porter, D. (2022). *Econometría*. Bogotá: McGraw-Hill.
- Hamzehi, M., & Hosseini, S. (2022). Business intelligence using machine learning. *Department of Computer Science, Faculty of Mathematics and Computer, Shahid Bahonar University*, 33233-33251.
- Helm1, J. M., Swiergosz1, A. M., Haeberle2, H. S., Karnuta1, J. M., Schaffer1, J. L., Spitzer3, A. I., . . . Krebs, V. E. (2020). Machine Learning and Artificial Intelligence: Definitions, Applications, and Future Directions. *Current Reviews in Musculoskeletal Medicine*, 69-76.
- Heras, J. M. (2020). Las 7 Fases del Proceso de Machine Learning - IArtificial.net. *IArtificial.net*. <https://www.iartificial.net/fases-del-proceso-de-machine-learning/>
- Hernández Sampieri, R., Fernández Collado, C., & Baptista Lucio, P. (2014). Metodología de la investigación. México: McGraw-Hill.
- Hotz, N. (2023). What is CRISP DM? - Data Science Process Alliance. Data Science Process Alliance. <https://www.datascience-pm.com/crisp-dm-2/>
- Huang, K. J. (2022). DeepValue: A Comparable Framework for Value-Based Strategy by Machine Learning. *Computational Economics*, 325-346.
- IBM. (2021). IBM Corporation. <https://www.ibm.com/docs/es/spss-modeler/saas?topic=nodes-statistical-models>
- IBM. What is Random Forest? (s.f.). <https://www.ibm.com/topics/random-forest>
- Implementation of Artificial Intelligence (AI): A Roadmap for Business Model Innovation. (2020). *MDPI*, 180-191.
- Indeed (2023) Sueldos- Indeed.com. <https://co.indeed.com/career/data-scientist/salaries>
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). An introduction to statistical learning. Springer.

- Jiang, Z., Xu, Y., Mao, H., & Jin, S. (2017). A deep reinforcement learning framework for the financial portfolio management problem. *IEEE Access*, 5, 24084-24094. doi: 10.1109/ACCESS.2017.2764845
- Joyanes, L. (2015). *Sistemas de información en la empresa*. Alfaomega.
- Joyanes, L. (2019). *Inteligencia de negocios y analítica de datos*. Alfaomega.
- Jurafsky, D., & Martin, J. H. (2020). *Speech and language processing. An introduction to natural language processing, computational linguistics, and speech recognition*. Pearson.
- Kimball, R. & Ross, M. (2013). *The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling* (3rd ed.). Wiley.
- Kreiger, J. R. (2020). Evaluating a Random Forest model - Analytics Vidhya - Medium. Analytics Vidhya. <https://medium.com/analytics-vidhya/evaluating-a-random-forest-model-9d165595ad56>
- Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*. https://proceedings.neurips.cc/paper_files/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf
- Kyan, M. (2014). *Unsupervised Learning*. In *John Wiley & Sons, Inc. eBooks*. <https://doi.org/10.1002/9781118875568>
- Laudon, K. C., & Laudon, J. P. (2018). *Sistemas de información gerencial* (15a ed.). Pearson
- Mauro, A. D., Sestino, A., & Bacconi, A. (2022). Machine learning and artificial intelligence use in marketing: a general taxonomy. *Italian Journal of Marketing*, 439-457.
- Milojević, N., & Redzepagic, S. (2021). Prospects of Artificial Intelligence and Machine Learning Application in Banking Risk Management. *Journal of Central Banking Theory and Practice*, 41-57.
- Modelo de madurez de operaciones de Machine Learning - Azure Architecture Center. (s.f.). Microsoft Learn: Build skills that open doors in your career. <https://learn.microsoft.com/es-es/azure/architecture/example-scenario/mlops/mlops-maturity-model>
- Montgomery, D. C., Peck, E. A., & Vining, G. G. (2022). *Análisis de regresión lineal*. John Wiley & Sons

- Morales, R. P., Villegas, C. M., & Zepeda, V. V. (2015). Análisis comparativo de modelos de madurez en inteligencia de negocio. *Ingeniare. Revista chilena de ingeniería*.
- Orlova, E. V. (2020). Decision-Making Techniques for Credit Resource Management Using Machine Learning and Optimization. *Department of Economics and Management, Ufa State Aviation Technical University*, 2-17.
- Pinelis, M., & Ruppert, D. (2022). Machine learning portfolio allocation. *The Journal of Finance and Data Science*, 35-54.
- Porter. (1980). *Strategy competitive*. Nueva York: Free Press.
- Ransbotham, S., Gerbert, P., Reeves, M., Kiron, D., & Spira, M. (2018). Artificial intelligence in business gets real. MIT sloan management review.
- Raschka, S., & Mirjalili, V. (2019). Python machine learning: Machine learning and deep learning with Python, scikit-learn, and TensorFlow 2 (3ra ed.). Packt Publishing.
- Reim, W., & Eriksson, J. Å. (2020). Implementation of Artificial Intelligence (AI): A Roadmap for Business Model Innovation. *PDMI*, 181-191.
- Roldán, R. C., Martín, E. M., & Rodríguez, A. R. (2019). *BIG DATA CON PYTHON Recolección, almacenamiento y proceso*. Bogotá: Alfaomega.
- Russell, S. J., & Norvig, P. (2016). *Artificial Intelligence: A Modern Approach* (3rd ed.). Pearson.
- Schneider, E., & Sabogal, C. (s/f). Modelo de optimización para la producción y comercialización de productos agrícolas en Colombia. <https://repositorio.uniandes.edu.co/server/api/core/bitstreams/903cfad4-7471-4b54-9746-f8be9471daad/content>
- Sethi, A. (2020). Support Vector Regression tutorial for machine learning. Analytics Vidhya. <https://www.analyticsvidhya.com/blog/2020/03/support-vector-regression-tutorial-for-machine-learning>
- Sklearn.Ensemble.GradientBoostingClassifier. (s/f). Scikit-Learn. <https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.GradientBoostingClassifier.html>
- Superintendencia de Subsidio Familiar. (2021). Informe Supersubsidio IA <https://www.ssf.gov.co/documents/20127/986361/INFORME+SUPERSUBSIDIO+IA.pdf/586ebbb8-401f-564e-06da-5f7b86c75870>

- Sutton, R. S., & Barto, A. G. (2018). Reinforcement learning: An introduction. MIT press.
- Talent. Salario para Científico Datos en Colombia - Salario Medio. (2023). Talent.com. <https://co.talent.com/salary?job=cient%C3%ADfico+datos>
- Talluri, S., & van Ryzin, G. (2004). The theory and practice of revenue management. Springer.
- Transforming Data With Intelligence (n.d). <https://tdwi.org/pages/about-tdwi/tdwi-business-intelligence-and-data-warehousing-education-and-research.aspx>
- Valenzuela, L., & Torres, E. (2008). GESTIÓN EMPRESARIAL ORIENTADA AL VALOR DEL CLIENTE COMO FUENTE DE VENTAJA COMPETITIVA. PROPUESTA DE UN MODELO EXPLICATIVO. Estudios Gerenciales.
- Vapnik, V. N. (1995). The nature of statistical learning theory. Springer
- Verdhan, V. (2020). Supervised Learning with Python. In *Apress eBooks*. <https://doi.org/10.1007/978-1-4842-6156-9>
- Wheelen, T., y Hunger, J. (2013). Administración estratégica y política de negocios. Pearson. Disponible en base de datos Ebook 7/24.
- Witten, I. H., Frank, E., & Hall, M. A. (2016). Data mining: Practical machine learning tools and techniques. Morgan Kaufmann Publishers.
- Wooldridge, J. M. (2021). Introducción a la econometría: Un enfoque moderno. Cengage Learning