

## Improving Opportunity Selection with Machine Learning to increase Hit Rate

Pedro Spencer Uriarte

Universidad Adolfo Ibañez  
ING500: PASANTÍA FULL-TIME  
Gonzalo Anriquez Gundian

24 de diciembre de 2023

## 1. Resumen

Los procesos de toma de decisiones de Bid no Bid y Go no Go son primordiales para la selección de oportunidades de negocios dentro de una empresa, estos procesos permiten descartar o seleccionar oportunidades en las cuales se invierte gran cantidad de recursos como tiempo o dinero. En el siguiente informe se propone una solución a la problemática de una toma de decisiones desinformada y basada en variables subjetivas sin el respaldo de datos históricos dentro del área de ventas de Grid Automation de la empresa Hitachi Energy, toma de decisiones que actualmente genera un bajo Hit Rate en distintos tipos de proyectos. La solución propuesta en el informe consiste en un modelo predictivo de clasificación binaria que entrega la probabilidad de éxito o de fracaso de una oportunidad de negocio junto con la influencia de cada una de las variables dentro de la predicción. En el informe se comparan distintos algoritmos de aprendizaje automático evaluando su desempeño, se utilizaron datos históricos de Brasil de los últimos cuatro años para su creación y evaluación. Los algoritmos utilizados fueron; Random Forest, AdaBoost, SVM, Decision Tree, Kneighbors y Redes Neuronales, con los cuales se realizó validación cruzada y prueba de distintos parámetros para ser ajustados a los requerimientos de la empresa. Una vez realizado el estudio se concluye que el algoritmo de Random Forest es el que tiene mejor desempeño con nuestros datos y se ajusta de mejor manera para lograr tener mayor certeza a la hora de tomar decisiones de Go no Go y Bid no Bid. Se realiza la implementación del modelo en tiempo real y finalmente se mide el desempeño de octubre y noviembre del año 2023, obteniendo un aumento del Hit rate y del Forecast accuracy en comparación con meses anteriores.

## Abstract

The decision-making processes of “Bid no Bid” and “Go no Go” are crucial for selecting business opportunities within a company. These processes enable us to discard or pursue opportunities where significant resources, such as time and money, are invested. In this report, we propose a solution to the problem of uninformed decision-making, which is currently based on subjective variables without the support of historical data within the sales area of Grid Automation at Hitachi Energy. This decision-making approach currently results in a low success rate for various types of projects.

The solution proposed in this report is a binary classification predictive model that provides the probability of success or failure of a business opportunity, along with the influence of each variable within the prediction. In the report, we compare different machine learning algorithms and evaluate their performance. We used historical data from Brazil from the last four years for its creation and evaluation. The algorithms used were Random Forest, AdaBoost, SVM, Decision Tree, K-neighbors, and Neural Networks. Cross-validation was performed with these algorithms, and different parameters were tested to align with the company’s requirements.

After conducting the study, we concluded that the Random Forest algorithm performs best with our data and is better suited to provide more certainty when making Go no Go and Bid no Bid decisions. The model was implemented in real-time, and its performance was measured in October and November 2023, resulting in an increase in the Hit rate and Forecast accuracy compared to previous months.

## **2. Contexto:**

### **2.1 Introducción:**

Hitachi Energy es una empresa global líder en tecnologías energéticas que ofrece soluciones y servicios en toda la cadena de valor de la energía, incluyendo soluciones como energía limpia, sistemas de generación de energía y soluciones de red eléctrica. Actualmente dentro de la empresa, específicamente en el área de ventas de Grid Automation existe una deficiencia en la toma de decisiones de los procesos “Go no Go” y “Bid no Bid”. Hoy en día estas decisiones se basan mayoritariamente en variables subjetivas, variables que muchas veces no tienen un respaldo concreto y no cuentan con información segura la cual les permitan tomar una decisión de manera más informada. Esta toma de decisiones desinformada se traduce en un bajo Hit Rate y bajo forecast accuracy dejando en evidencia el gran número de oportunidades ingresadas que no son exitosas.

### **2.2 ¿Para Qué?**

El propósito fundamental del proyecto es lograr una mejora en la calidad de las decisiones tomadas en cuanto a la selección de las oportunidades, esto con el propósito de contar con oportunidades las cuales tengan una mayor probabilidad de éxito, al disponer con oportunidades de mayor calidad se estaría generando una reducción de riesgos asociados.

### **2.3 Dolores**

La empresa actualmente se ve altamente expuesta a realizar propuestas invirtiendo tiempo y recursos en oportunidades que en base a datos históricos podrían tener una baja probabilidad de éxito o a su vez se podrían estar perdiendo oportunidades las cuales tienen alta probabilidad de éxito, pero no se están tomando en cuenta. Lo anterior genera una falta de confiabilidad y precisión en la selección de las oportunidades, evidenciándose esta problemática en un bajo Hit Rate en diversas categorías de las oportunidades, además se presenta un bajo forecast accuracy, indicando que oportunidades de las cuales se cree tener alta probabilidad de éxito, ya que cumple con parámetros de confiabilidad, no están teniendo el éxito esperado.

### **2.4 Oportunidad de mejora**

Se detecta una importante oportunidad de mejora en el área de toma de decisiones y selección de proyectos, es necesario aplicar herramientas que permitan disminuir riesgos y entregar mayor seguridad a la hora de seleccionar oportunidades, esto permitirá tener mayor certeza del resultado que tendrán las oportunidades que ingresen a la empresa, esperando que se refleje en un aumento del Hit Rate y un ahorro tanto de tiempo como de recursos invertidos.

Existen distintos datos que demuestran que hoy en día se está generando esta problemática, Hit Rates Bajos en algunas categorías tales como:

- Línea de producto 2875: 21% Hit Rate
- Rangos de precio: 200.000 \$ – 300.000 \$ Hit Rate de 25%
- Opportunity Owner: Vendedores con 5% de Hit Rate
- Account Customer: Clientes con 9% de Hit Rate

Para evidenciar el bajo rendimiento en estas categorías se puede apreciar la diferencia con otras como:

- Línea de producto 2879: 80% Hit Rate
- Rangos de precio: 150.000 \$ – 200.000 \$ Hit Rate de 51%
- Opportunity Owner: Vendedores con Hit Rates sobre 70%
- Account Customer: Clientes con Hit Rates sobre 70% Hit Rates y con la misma cantidad de oportunidades ingresadas.

Estos comportamientos se fueron descubriendo durante el análisis exploratorio de datos, Hit Rates los cuales se presentan como bajos, en comparación con las categorías similares. Estos comportamientos afectan directamente al Hit Rate de la empresa y de los cuales no se tenía conocimiento.

### 3. Objetivos:

#### 3.1 Objetivo General

Lograr mediante una herramienta de apoyo en la toma de decisiones de Bid no Bid o Go no Go para los proyectos, una mejora en el proceso de selección de oportunidades, aumentando la tasa de éxito medida por el Hit Rate en 10% y medido en un plazo de 2 meses.

#### 3.2 Objetivos Específicos

- **Hit Rate:** Lograr aumentar el Hit rate en un 10% en base a oportunidades ganadas vs total de oportunidades de Brasil, para luego medir este aumento durante el mes de noviembre.
- **Forecast Accuracy:** Generar un aumento en la tasa de éxito de oportunidades que cuentan con mayores parámetros de confiabilidad, las cuales se encuentran en el forecast accuracy. Se espera poder aumentarlo a un 85% durante el mes de noviembre.
- Lograr encontrar al menos 5 comportamientos que puedan ser útiles para la toma de decisiones y que permitan disminuir algunos riesgos asociados y entreguen más información a la toma de decisiones, mediante la importancia de cada característica en el resultado de cada oportunidad.

#### 4. Estado del Arte

La problemática de toma de decisiones en procesos de Bid no Bid y Go no Go, es una problemática que afecta a un gran número de empresas en diferentes sectores, ya sea la Minería, Tecnologías, empresas financieras o prácticamente todas las empresas que se ven enfrentadas a tomar decisiones estratégicas sobre la asignación de los recursos para poder reducir los costos de horas/hombre, costos administrativos, costos de oportunidad o diversos costos que tiene asociados el realizar ofertas o nuevos proyectos.

Muchas empresas se ven afectadas a la hora de tomar decisiones, ya sea porque no cuentan con datos concretos, no realizan análisis de riesgos, utilizan variables subjetivas o principalmente porque no se definen de manera correcta los distintos parámetros que debe cumplir una nueva oportunidad para poder llevarse a cabo, generando que existan muchos proyectos que fracasan, por consecuencia un Win Rate Bajo y un gasto innecesario de recursos invertidos.

Lo que explica la literatura actualmente con respecto a la toma de decisiones de Bid no Bid principalmente en empresas de construcción y que participan en procesos de licitación, es que se proponen modelos de decisión y diagramas de flujo los cuales deben cumplir múltiples criterios o factores que afectan a los procesos de licitación, tal como nos indica el Artículo “*DECISION-SUPPORT SYSTEM FOR MODELING BID/NO-BID DECISION PROBLEM*”, “*Factors such as type of job, owner, location, size of job, current work load, and strength of the firm are important for the bid/no-bid decision; degree of hazard, degree of difficulty, uncertainty in the estimate, risk in investment, and reliability of subcontractors are important for the percent markup decisions*”(Ahmed 1990), dándonos a entender que para apoyar la toma de decisiones es muy importante tener en cuenta diversos factores que pueden aumentar la confiabilidad de las oportunidades. Los principales criterios que se proponen en gran número de artículos relacionados con esta problemática son los siguientes:

1. Objetivos estratégicos de la empresa y factores de atractivo
2. Requisitos mínimos
3. Identificación de la competencia
4. Relación con el cliente
5. Plazo para ejecutar el proyecto
6. Riesgos asociados

Si bien los distintos artículos proponen algunos criterios más o algunos criterios menos, todos presentan modelos de tomas de decisiones que deben cumplir con criterios mínimos para tomar la decisión de si ofertar o no.

Como ejemplo, el artículo (Araujo, 2021)” (De Araújo et al., 2021) propone el siguiente modelo para la toma de decisiones:

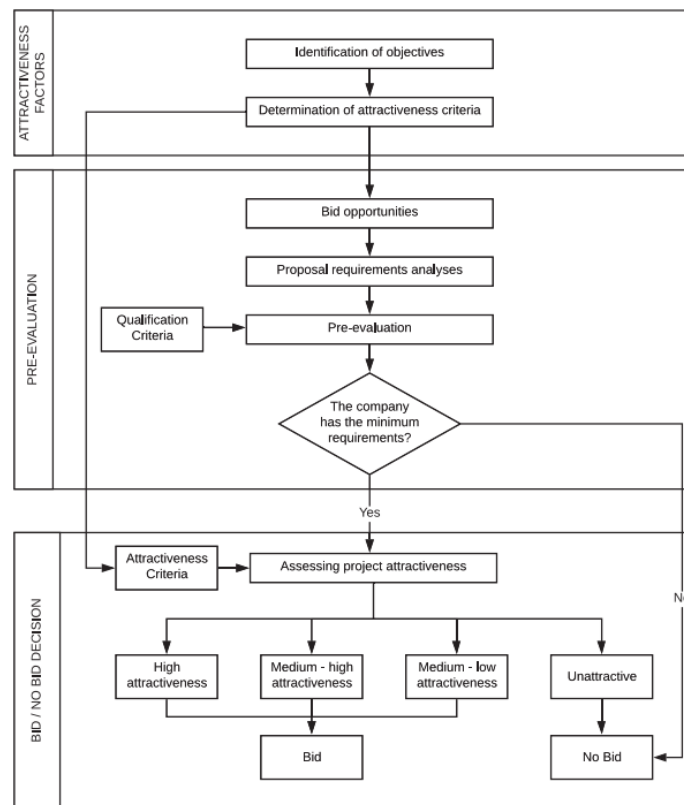


Fig. 1. Projects' classification model for bidding decision.

Figura 1: Modelo de decisión (Araujo, 2021)

Como se puede apreciar el modelo cuenta con criterios como Factores de atractivo o requisitos mínimos de preevaluación.

Actualmente la empresa Hitachi Energy ya cuenta con un proceso de toma de decisiones que tiene en cuenta estos factores, que tal como se expuso en la problemática, estos factores pueden llegar a ser factores subjetivos, la empresa actualmente cuenta con criterios como; objetivos estratégicos, tipo de oportunidad, presupuesto, relación con el cliente, plazo de la oportunidad, análisis de riesgos asociados, identificación de la competencia y gran parte de lo que expone la literatura sobre esta problemática ya está implementada en la toma de decisiones de la empresa. Pero una de las soluciones propuestas por la literatura actual y más reciente es el uso de ciencia de datos para la toma de decisiones, debido a que actualmente la toma de decisiones se basa en muchas variables subjetivas, podemos ver que la literatura presenta lo siguiente; “debido al alto grado de incertidumbre que puede tener una decisión, el data science presenta un rol muy importante, debido a que el análisis exhaustivo de data y la generación de conocimiento, permiten predecir de manera más acertada posibles escenarios o resultados futuros, con lo cual se reduce el riesgo de la decisión tomada. En consecuencia, la utilización del data science en la toma de decisiones genera un valor añadido.” (Pignano y Pino, 2021). Dentro del data Science Podemos encontrar la minería de datos y los modelos de machine Learning, los cuales son utilizados muchas veces para apoyar la toma

de decisiones, pudiendo descubrir patrones y encontrar comportamientos que nos entreguen mayor información antes de tomar una decisión como la es el Bid/no Bid, se expone que *“las herramientas de apoyo a la toma de decisiones podrían utilizarse también como herramientas para eliminar los resultados innecesarios e irrelevantes obtenidos de la Minería de Datos.”* (Marcano y Talavera, 2007) permitiéndonos así mediante a un modelo de Machine Learning eliminar resultados innecesarios o descartar oportunidades las cuales basado en los datos no tendrán el resultado deseado por la empresa.

Dado que la empresa Actualmente cuenta con lo que la literatura propone como la mejor práctica que es los modelos de decisiones y aplicación de múltiples criterios para poder apoyar la toma de decisiones, es importante poder apoyar este modelo con el uso de Data Science y Machine Learning y así poder anticiparnos a los resultados que tendrá cada una de las oportunidades, pudiendo descartar algunas y seleccionar las que cuentan con una mayor probabilidad de éxito.

Una de las primeras etapas de la toma de decisiones en Hitachi Energy, específicamente para el área “2851”, es la siguiente:

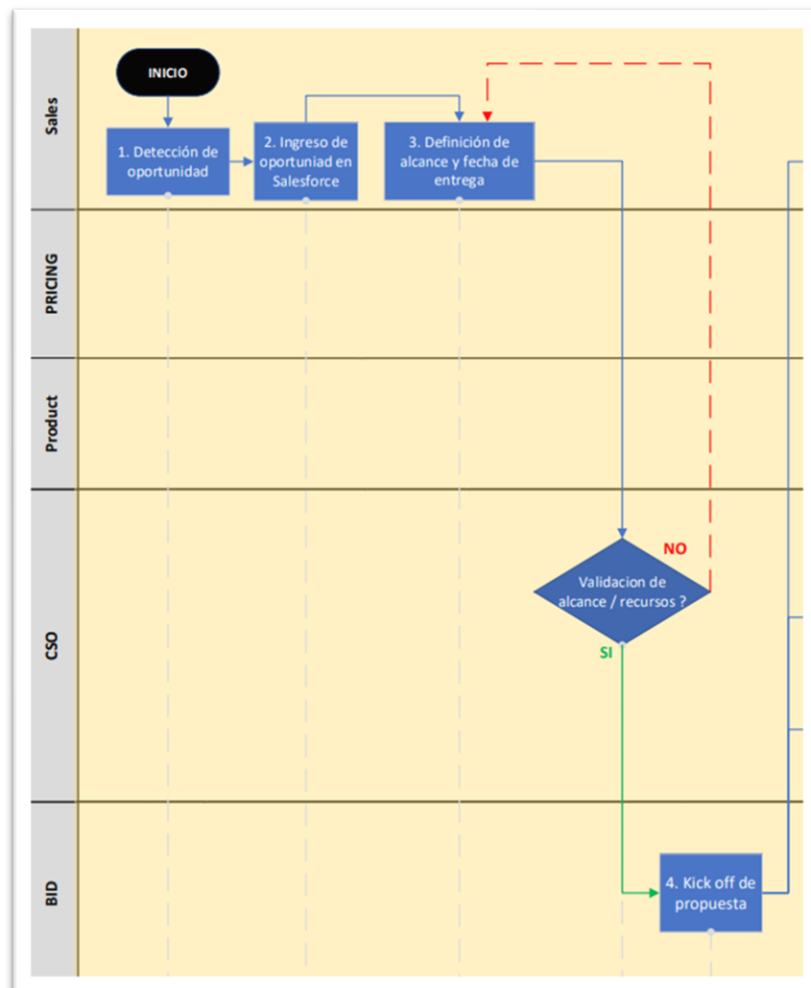


Figura 2: Primera etapa modelo decisión Hitachi Energy “2851”



Esta primera etapa debe cumplir con los siguientes criterios para poder ir avanzando en el diagrama:

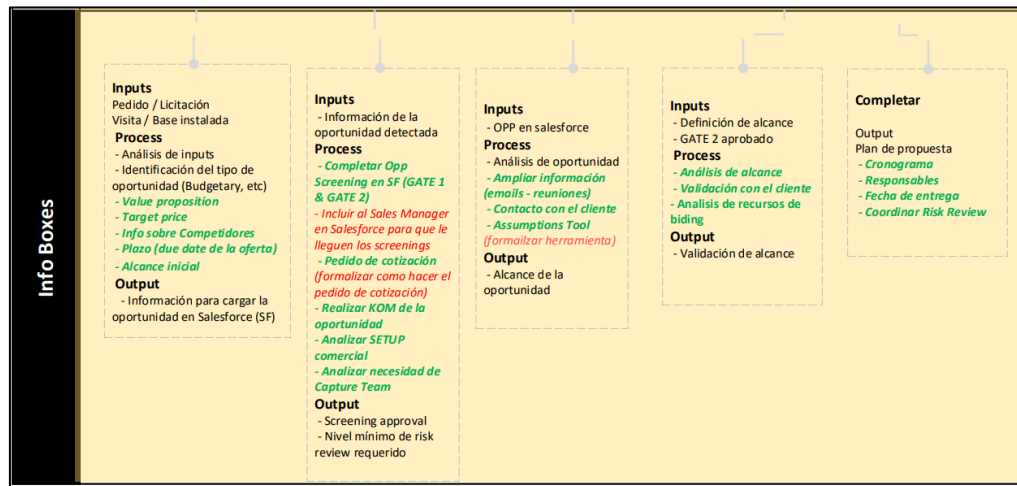


Figura 3: factores y criterios toma de decisiones Hitachi Energy “2851”

Como se puede apreciar, existen diversos criterios o pasos que la oportunidad debe ir cumpliendo para poder avanzar con la toma de decisiones. Mediante un modelo predictivo que utilice los datos históricos de todas las oportunidades, que nos permita predecir si la oportunidad será “Won” o “Lost”, se podría descartar un gran número de oportunidades en las cuales se invierte dinero y también se podría poner mayor énfasis en oportunidades las cuales tienen una mayor probabilidad de ser exitosas según el modelo predictivo.

Es importante destacar que esta etapa inicial es una etapa que conlleva una gran inversión de tiempo y por consecuencia de recursos, la etapa “Pursuit” etapa en donde se toma la decisión de “Bid no Bid” en promedio toma alrededor de 1 a 2 meses de trabajo de revisión de estándares de confiabilidad e investigación de la oportunidad.

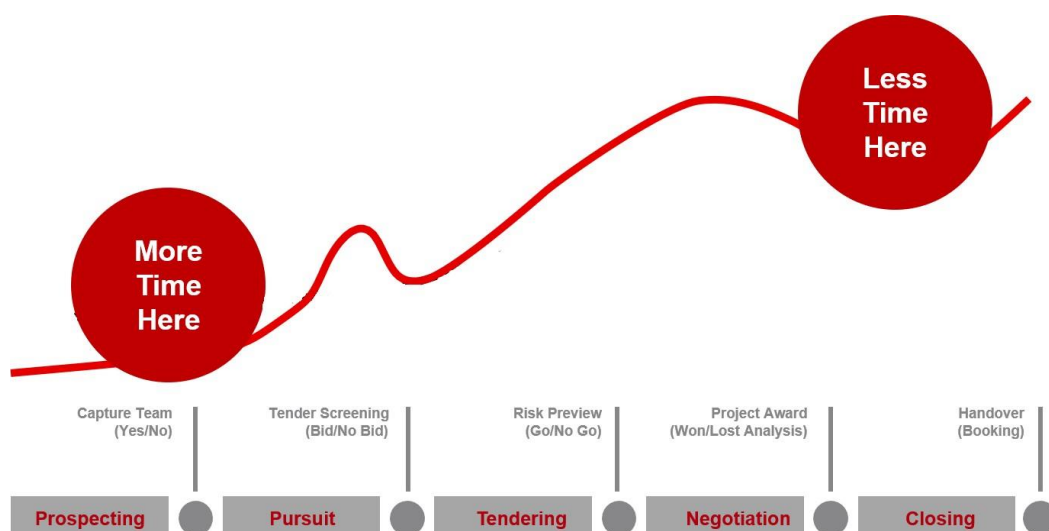


Figura 4: Aumento de tiempo invertido en cada etapa de una oportunidad en Hitachi Energy

Es importante que la empresa seleccione de manera correcta las oportunidades para evitar invertir tiempo y recursos en oportunidades que finalmente no son exitosas.

Para poder realizar el modelo predictivo es necesario usar un modelo de aprendizaje supervisado y de clasificación Binaria, ya que la idea es poder predecir el resultado de la oportunidad y saber si el resultado será “Won” o “Lost”.

Dentro de los modelos de machine Learning podemos encontrar una gran variedad que se presenta en las mejores prácticas para modelos de clasificación, de los cuales podemos encontrar los siguientes:

- **Decision Tree:** Es uno de los modelos más utilizados para problemas de clasificación, la literatura explica que; *“Los árboles de decisión crean un modelo de clasificación basado en diagramas de flujo. Clasifican casos en grupos o pronostican valores de una variable dependiente (criterio) basada en valores de variables independientes (predictoras).”* (Berlanga, V., Rubio Hurtado, M. J., & Vilà Baños, R. 2013).
- **K-Nearest Neighbor:** Principalmente usado para predecir la clase de un objeto en base a las características, la literatura indica que *“El método K-nn (K nearest neighbors Fix y Hodges, 1951) es un método de clasificación supervisada (Aprendizaje, estimación basada en un conjunto de entrenamiento y prototipos)”* (Parra, 2019)
- **Support Vector Machines:** También es utilizado para problemas de clasificación supervisada del cual se expone que: *“es un modelo que partiendo de un conjunto de ejemplos de entrenamiento, podemos etiquetarlos en diferentes clases y representar dichas muestras en puntos en el espacio para tratar de separar las diferentes clases”* (Parra, 2019)
- **Random Forest:** Al igual que los demás modelos, random forest presenta distintos usos; *“Random Forest ha demostrado ser efectivo en una amplia variedad de dominios y problemas, como la clasificación de imágenes (Cutler et al., 2012), el análisis de datos genéticos (Bureau et al., 2005), la detección de fraudes (Bolón-Canedo et al., 2014) y la ecología (Prasad et al., 2006).”* (Alaminos-Fernández, A. F. (2022).
- **Redes Neuronales MLP:** Las redes neuronales tienen distintos usos dentro de las predicciones y existen distintos tipos de redes neuronales, uno de los principales usos es para problemas de clasificación y se expone que; *“El modelo más utilizado es el perceptrón multicapa, que abarca el 70% de las aplicaciones, dado que se ha demostrado que este modelo es un aproximador universal de funciones (Funahashi 1989)”* (Parra, 2019). Si bien se descarta en una primera etapa el uso de redes neuronales debido a que su uso generalmente es para problemas con una gran cantidad de datos y la data de Hitachi Energy en Brasil no cuenta con suficientes datos, finalmente se llega a la conclusión de probar su desempeño y compararlo con los otros modelos antes de descartarlo.

La literatura nos expone una gran variedad de soluciones en cuanto a modelos predictivos para problemas de clasificación, si bien todos estos modelos se adaptan a nuestra necesidad, es importante medir su desempeño con los datos que presenta la empresa para así evaluar cuál es el indicado para este proyecto específico.

## 5. Solución

Según lo estudiado, la empresa Hitachi Energy cuenta con las mejores prácticas que se proponen en la literatura, dado esto es importante comprender las causas de por qué las oportunidades están fallando antes de modificar directamente el modelo de decisión de la empresa. Debido a lo anterior se escoge la solución de la creación de un modelo predictivo con el uso de machine Learning. Actualmente la empresa cuenta con 3 resultados de las oportunidades “Won”, “Lost” y “Closed Lost Without Bid”, se evaluó la creación de un modelo de clasificación multiclase, pero la empresa no cuenta con suficientes datos como para generar un modelo preciso con 3 etiquetas, es por esto por lo que se llega a la conclusión de realizar un modelo de aprendizaje supervisado de clasificación Binaria. Se incluyeron las oportunidades “Closed Lost Without Bid” a las oportunidades “Lost”, ya que si bien son un resultado distinto porque no se realiza la oferta y no se invierten tantos recursos, aun así, se invierte tiempo en ingresar las oportunidades y cuentan con un comportamiento similar a “Lost”, por lo que es importante poder identificar este tipo de oportunidades y descartarlas lo antes posible para generar ahorro en tiempo y recursos invertidos.

Sumado a lo anterior es importante destacar que se realizó el análisis de datos correspondiente, limpieza de los datos, análisis descriptivo y exploratorio, esto fue útil para poder seleccionar 12 variables las cuales se usarán para probar los distintos algoritmos. Las variables seleccionadas fueron variables que deben cumplir las oportunidades antes de comenzar con el proceso de toma de decisiones, esto para poder anticiparnos en una etapa temprana al resultado que tendrá la oportunidad.

En una primera Etapa para la selección de las variables a utilizar se realizó el proceso de Análisis de datos, en el cual se llevaron a cabo las siguientes actividades:

1. Análisis Univariado
2. Análisis Bivariado
3. Análisis Multivariado
4. MCA (Análisis de correspondencia Múltiple)
5. Análisis de Chi cuadrado
6. Análisis de Clúster

Esto se realizó con el propósito de entender de manera general como se relacionan las variables con el resultado de una oportunidad y así descartar variables las cuales no cuentan con relaciones significativas.

Finalmente se probaron 6 distintas soluciones en cuanto a modelos de aprendizaje automático, los modelos que se evaluaron fueron:

- AdaBoost
- DecisionTree
- Kneighbors
- Random Forest
- Redes Neuronales MLP
- SVM

Para evaluar el desempeño de cada uno se realizó la búsqueda de los mejores parámetros mediante GridSearch y a su vez se aplicó validación cruzada para poder garantizar la exactitud de las predicciones del modelo y obtener un resultado confiable, utilizando 80% de los datos como datos de entrenamiento y 20% para probar el desempeño. A su vez se calculó el Recall, Accuracy y desviación estándar de cada modelo para poder comparar sus desempeños y seleccionar el que se adapta mejor a las necesidades de la empresa. Cada solución entregó el siguiente desempeño:

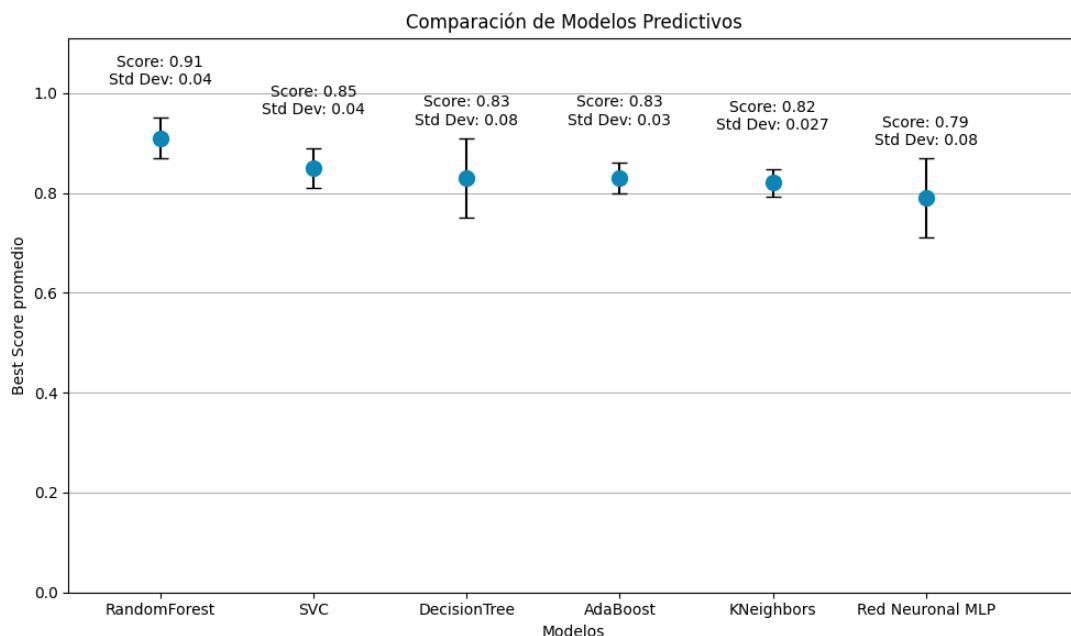


Figura 5: Comparación de Recall promedio de cada modelo

	Random Forest	SVM	DecisionTree	AdaBoost	Kneighbors	Redes Neuronales MLP
Cross validation Best Score	91%	85%	83%	83%	82%	79%
Desviación estandar	4%	4%	8%	3%	3%	8%
Accuracy	84%	82%	77%	78%	81%	80%

Figura 6: Tabla de comparación de Recall y Accuracy de cada modelo

Luego del estudio se llegó a la conclusión de que la solución que se adapta de mejor manera a los datos de la empresa es el modelo de Random Forest, el cual logró tener un Score promedio de 91%, calculado por el Recall, medida de desempeño que es seleccionada debido a que se adapta a las necesidades de la empresa y permite enfocarnos en el aumento de oportunidades ganadas, ya que la empresa tiene como objetivo apuntar al mayor número de oportunidades que puedan ser “Won” y no le da mucha importancia si una oportunidad es perdida ya que el costo no es tan elevado en comparación con lo que se ganaría con una oportunidad. además, el modelo cuenta con una desviación estándar moderada, lo que entrega un grado de confianza en el resultado.

## 6. Evaluación Económica

La evaluación económica de este proyecto en específico no conlleva gran complejidad ni muchos costos para la empresa, la evaluación consta de 4 aspectos importantes a tener en cuenta:

- **Costo de desarrollo:** El proyecto tiene un desarrollo de más de 810 horas, actualmente la empresa cuenta con un apoyo económico a los pasantes de \$300.000 CLP mensuales, esto traducido a los 5 meses aproximadamente que dura el proyecto son \$1.500.000 CLP que incurre la empresa en costo de desarrollo del proyecto.
- **Costo de implementación:** Es importante tener en cuenta la plataforma en la que se realizará el modelo predictivo. Se propone una Web local que permita aplicar el modelo en tiempo real, Web que no tiene costos de implementación, por lo que el proyecto no generará un costo adicional para la empresa en cuanto a IDE.
- **Capacitación:** Es importante poder realizar una buena capacitación del personal que utilizará el modelo predictivo, esto genera un costo H/H invertido por el personal, si bien no es un costo adicional para la empresa, aproximadamente se invierten \$90 USD por cada H/H y es importante tenerlo en cuenta.
- **Costo de mantención y actualización:** A futuro es necesario que el modelo tenga un mantenimiento y una actualización de los datos y así poder ir adaptándolo a los cambios que se irán generando, eso requerirá de la contratación de un Data Scientist con el conocimiento para poder mantener el modelo, el cual será un costo adicional para la empresa.

En cuanto a los beneficios estimados que entregará el proyecto podemos encontrar los siguientes:

- **Ahorro en costos:** El participar en licitaciones y llevar a cabo todo el proceso de oferta, tiene un costo promedio de 30.000 \$, por lo que poder seleccionar de mejor manera las oportunidades generaría un ahorro significativo para la empresa.
- En promedio las oportunidades que entran a la empresa son de 200.000 \$ aproximadamente, además anualmente ingresan cerca de 100 oportunidades, por lo que generar un aumento del 10% en el Hit rate de la empresa, generaría ingresos extras de 2.000.000 \$

Esto evidencia en gran parte la rentabilidad que genera el realizar un proyecto como este, el cual no requiere de muchos costos adicionales para la empresa y genera grandes beneficios.

## 7. Metodología

Se selecciona la metodología CRISP-DM, ya que se adapta a las necesidades del proyecto, partiendo por el entendimiento de la estructura del negocio y la estructura de los datos para así poder comprender en términos generales la situación actual de la empresa, siguiente a esto encontrar comportamientos y patrones que indiquen las áreas específicas que tienen relación con el resultado de las oportunidades y así seleccionar de manera correcta las variables necesarias que se utilizarán en la solución. Además, la literatura expone que; *“En el caso de los proyectos de implementación de minería de datos una de las metodologías que ha tenido más apoyo de las empresas privadas y organismos públicos es CRISP-DM”* (Galán Cortina, 2016). Siendo esta metodología una de las mejores prácticas para realizar este tipo de proyectos.

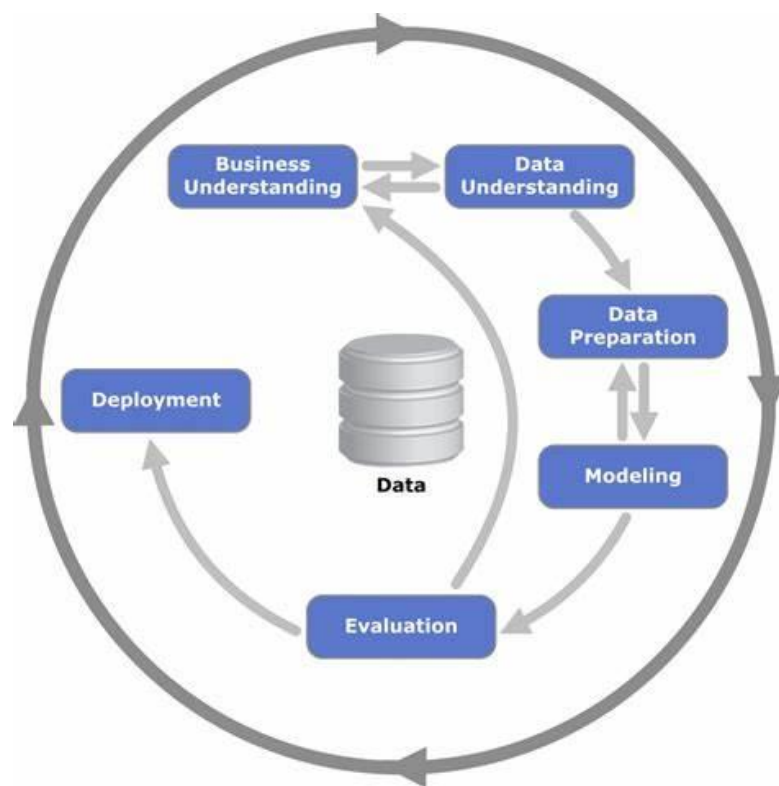


Figura 7: Pasos Metodología CRISP-DM

## 8. Plan de implementación

El plan de implementación se acopla de buena manera a la metodología seleccionada CRISP-DM, se propuso el siguiente plan de implementación para poder llevar a cabo la realización completa del proyecto:



- **Recopilación de datos:** Obtención de datos de Salesforce específicamente de Brasil de los años 2019 a 2023 con un presupuesto de la oportunidad entre 10.000 \$ – 1.200.000 \$.
- **Análisis de datos:** Análisis Univariado, Bivariado, Multivariado, MCA, Análisis de Chi cuadrado y análisis de Cluster con el fin de descartar y seleccionar las variables que se utilizaran en la creación del modelo.
- **Selección del Modelo:** Se Estudian las posibles soluciones y se define un modelo de clasificación binaria, el cual luego de realizar estudios y pruebas de cada uno de los modelos propuestos, se comparan los desempeños para seleccionar el más indicado con respecto a las métricas que buscamos.
- **Creación y Evaluación:** Entrenamiento del modelo con las variables seleccionadas, evaluar desempeño junto con la búsqueda de los mejores parámetros y validación cruzada para finalmente evaluar como se comporta con datos nuevos del 2023.
- **Capacitación:** Seleccionar el personal que utilizará y tendrá acceso al modelo, que se familiaricen y entiendan las limitaciones junto al debido uso que le deben dar al modelo.
- **Implementación:** Puesta en marcha en tiempo real y apoyo al personal en caso de ser necesario.
- **Medir desempeño:** Poder finalmente evaluar el desempeño medido por el forecast accuracy y el Hit Rate. Una vez ya se haya puesto en marcha el modelo.

## 9. Matriz de Riesgos

Riesgo	Impacto	Probabilidad	Prioridad	Mitigación
Baja calidad de los datos utilizados	Muy Alto (5)	Muy Bajo (1)	6	Selección correcta de los datos, gran numero de oportunidades y limpieza de datos
Baja precisión del modelo predictivo	Alto (4)	Alto (4)	8	Probar distintos modelos de machine learning y correcta selección de variables a utilizar
Alto tiempo de creación del modelo y adaptación a pagina Web	Alto (4)	Bajo (2)	6	Estudiar adaptacion de Modelo predictivo a Web Local para puesta en marcha
Bajo desempeño con datos reales	Muy Alto (5)	Bajo (2)	7	Correcto entrenamiento del modelo y realización de pruebas antes de implementación
Mal o Nulo uso del Modelo por parte del personal debido a desinformación	Alto (4)	Media (3)	7	Correcta capacitación y selección del personal que utilizará el modelo
Poco tiempo para medir desempeño del modelo	Alto (4)	Muy Alto (5)	9	implementar modelo en noviembre y medir oportunidades con bajo plazo

*Figura 8: Matriz de riesgos*

La Matriz de Riesgos expone los 6 principales riesgos que conlleva la realización de este proyecto, fue utilizada como guía para entregar prioridad a las distintas actividades que tienen un mayor impacto y una mayor probabilidad de suceder, para así generar las mitigaciones correspondientes y evitar los riesgos asociados al proyecto.

## 10. Medidas de Desempeño

Para poder validar el éxito o fracaso del proyecto es importante tener en cuenta las siguientes métricas que nos ayudaron a evaluar el desempeño de la solución.

- **Validación del modelo predictivo:** Específicamente para este proyecto es importante anticiparnos a cualquier resultado de la oportunidad, pero principalmente lo que nos importa es no dejar pasar oportunidades que sean “Won” y así aumentar el Hit Rate, debido a esto es necesario utilizar Recall como medida de validación del modelo predictivo; ( $\text{True Positives} / (\text{True Positives} + \text{False Negatives})$ ). En base al score de cada uno de los modelos propuestos se selecciona el que entregue un mejor resultado, siempre teniendo en cuenta la desviación estándar y el Accuracy, pero como medida principal el Recall.
- **Hit Rate:** Se espera poder aumentar el Hit rate de Brasil a un 50%, a su vez se espera poder comparar el Hit rate de octubre y noviembre del año 2022, el Hit rate de agosto y septiembre del año 2023 y finalmente una vez aplicado el modelo medir su desempeño en los meses de octubre y noviembre del año 2023.
- **Forecast Accuracy:** Se espera poder obtener un forecast accuracy sobre el 85% en el mes de octubre y noviembre, es decir, de las oportunidades que se encuentran en el forecast, oportunidades que cumplen con parámetros de confiabilidad deben tener un accuracy sobre el 85%.



## **11. Desarrollo e implementación de la solución**

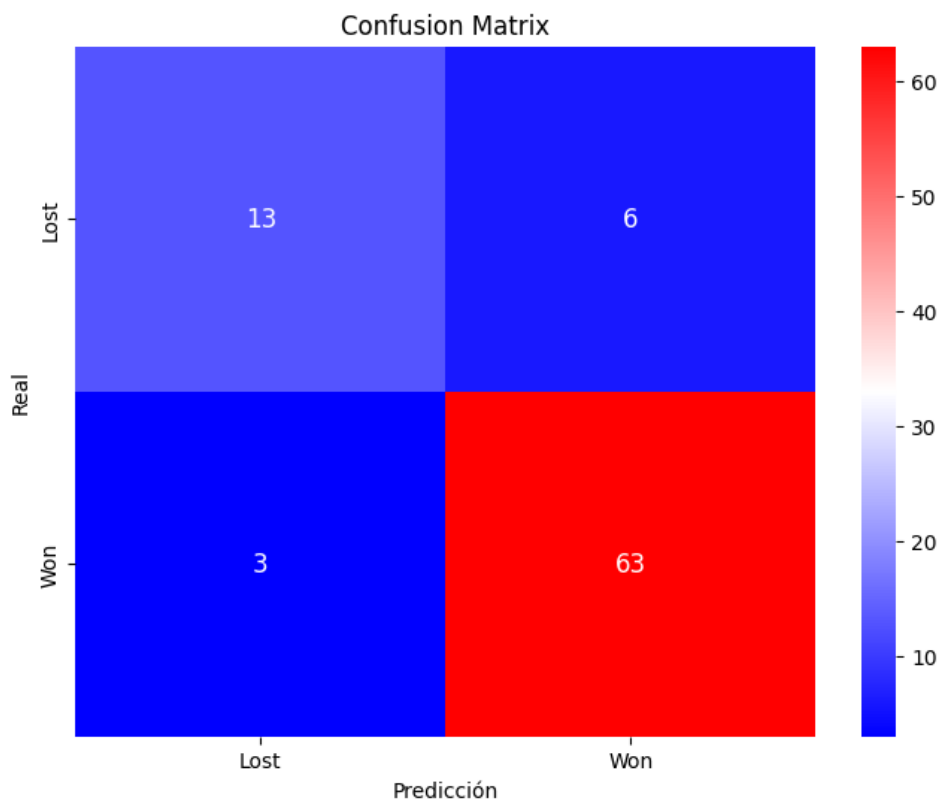
Una vez llevada a cabo la medición de desempeño de cada una de las soluciones propuestas, midiéndolas a través de sus mejores parámetros y validación cruzada, se selecciona el modelo de Random Forest, el cual obtiene un Best Score promedio de 91% y se guarda para posteriormente ser utilizado. Dado esto se comienza con el desarrollo del entorno en el que el personal de la empresa podrá utilizarlo. Se decide llevar a cabo el desarrollo del proyecto en una Web la cual permite a los usuarios seleccionar cada una de las variables, se selecciona este entorno debido a que no genera un costo para la empresa y es amigable para el personal.

Ya creada la web que permite al usuario seleccionar las variables e ingresar una nueva oportunidad, la página entrega el resultado de la predicción y la probabilidad, utilizando el modelo previamente entrenado, junto con esto se despliega un gráfico que muestra las variables más influyentes en el resultado de la predicción, este gráfico fue desarrollado mediante la herramienta Shap de Python que nos permite visualizar la importancia de cada una de las características que utiliza el modelo para crear la predicción, a su vez se probó la herramienta LIME y Feature importance para apreciar cuales eran las variables mas relevantes, pero se decide usar Shap ya que nos permite ver la importancia de cada característica en la predicción específica, además de ser una herramienta fácil de interpretar y de manejar.

Finalmente se comienza con la capacitación del personal, enseñándole las limitaciones y el correcto uso del modelo, una vez finalizada la capacitación se pone en marcha el modelo, comenzando la prueba con 8 oportunidades que fueron ingresando a la empresa durante los meses de octubre y noviembre para posteriormente analizar sus resultados y el desempeño del modelo.

## 12. Resultados

Luego de haber seleccionado el modelo de random forest y los parámetros correspondientes los cuales entregaron un mayor desempeño en la validación cruzada, se probó el desempeño utilizando oportunidades nuevas del año 2023, es decir oportunidades las cuales no se habían utilizado para entrenar el modelo, el resultado fue el siguiente:



*Figura 9: Matriz confusión oportunidades nuevas 2023*

Como se puede apreciar en la imagen, el resultado del modelo fue bastante acertado, entregándonos un accuracy de 89% y un recall de 95%. La matriz confusión nos entrega 3 principales conclusiones de los resultados;

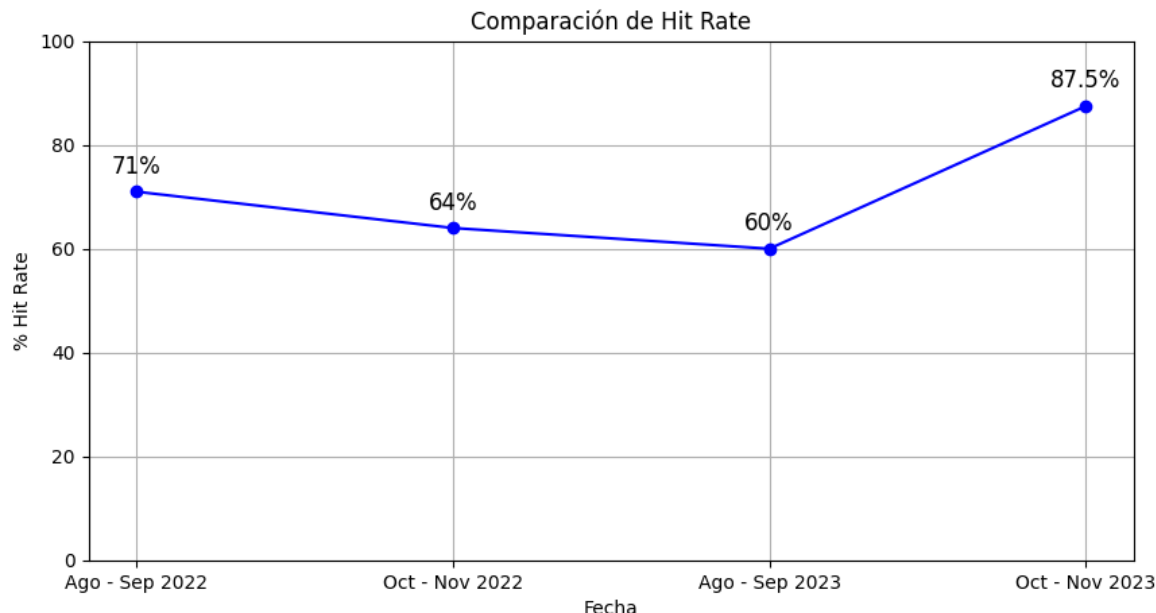
- Del total de oportunidades que fueron ganadas por la empresa, el modelo fue capaz de predecir 63 de 66, entregándonos un recall de 95%
- Del total de oportunidades perdidas el modelo fue capaz de predecir 13 de 19, este resultado es esperado, ya que el modelo fue entrenado enfocado en la métrica Recall.
- Finalmente, del Total de oportunidades, el modelo acertó 76 oportunidades de 85, equivocándose solo en 3 oportunidades que fueron “Won” las cuales son las que nos interesan para el aumento del Hit rate.

Una vez pudimos comprobar que efectivamente el modelo estaba realizando un buen trabajo con la predicción de las oportunidades, se comenzó la prueba en tiempo real de nuevas oportunidades ingresadas a la empresa.

Durante el mes de octubre y noviembre ingresaron a la empresa un total de 8 oportunidades las cuales tenían como fecha esperada de cierre antes del mes de diciembre, se llevó a cabo la prueba de cada una de las oportunidades y el modelo indicó que las 8 oportunidades serían ganadas. Una vez utilizado el modelo se apoyó la toma de decisiones de Bid no Bid y Go no Go para dar las indicaciones correspondientes según lo que el modelo indicó, es decir que en todas las oportunidades se debía ofertar y realizar el “Go” ya sea del proyecto o la venta del producto. Durante el Mes de noviembre se comenzó con el cierre de cada una de las oportunidades que se habían abordado, y el resultado fue el siguiente:

- Del total de las 8 oportunidades que ingresaron a la empresa, el modelo fue capaz de predecir 7 de manera correcta, es decir se logró un Hit Rate del 87.5%.
- La oportunidad la cual no fue acertada por el modelo fue una oportunidad cancelada por el cliente, es decir que no fue una oportunidad en la cual se ofertó ni se tomó alguna decisión de Bid no Bid o Go no Go. Como resultado, el forecast accuracy fue de un 100%, por lo tanto, del 100% de las oportunidades que se encontraban en el forecast y se llevó a cabo la decisión de Go no Go fueron todas ganadas por la empresa.

Para poder medir el aumento que se generó con respecto a los meses anteriores se calculó el hit rate de 3 periodos con proyectos similares, es decir proyectos que tuvieran una duración esperada de máximo 2 meses y que se encuentren dentro del rango de precios de 0 a 200.000 \$, para luego comparar la diferencia con el periodo en el cual se utilizó el modelo creado.



*Figura 10: Comparación Hit Rates por periodos*

Como se puede apreciar en el gráfico, existió un aumento del hit rate en promedio sobre 20%, siendo este un buen resultado y cumpliendo con el objetivo de generar un aumento del hit rate un 10%, a su vez cumple con el objetivo de mantener un forecast accuracy sobre el 85%, ya que se obtuvo un forecast accuracy del 100%.

Hay que destacar que si bien el número de oportunidades con las que se probó en tiempo real el modelo es bajo, el resultado que entregó el modelo fue mejor de lo esperado, además se realizaron pruebas con meses anteriores para probar que predicción entregaba a oportunidades que ya se habían concretado y tuvo una precisión por sobre el 90% en todas las pruebas que se realizaron, midiendo el desempeño en el mes de junio, julio, agosto y septiembre se lograron predecir 45 oportunidades de un total de 50.

También es importante mencionar que las 8 predicciones realizadas por el modelo tuvieron como resultado “Won” con una probabilidad todas sobre 80% entregando cierto grado de confianza en las decisiones que se tomaron.

Es necesario agregar que no se presentan resultados en base a aumentos monetarios debido a que los ingresos de la empresa dependen altamente de las oportunidades que se presentan y esto puede variar en todos los meses, pudiendo así en algunos meses ingresar oportunidades sobre 2.000.000 \$ y otros meses solo oportunidades de 100.000 \$ es por esto que el resultado se mide en el aumento de Hit Rate y no basándonos en ingresos económicos, pero dado que se estima el ingreso promedio de las oportunidades en 200.000 \$, el generar un aumento del 10% en el Hit Rate generaría un aumento de 2.000.000 \$, anuales ya que en promedio ingresan 100 oportunidades al año.

Se logró también encontrar distintos comportamientos que afectan a los resultados de cada oportunidad mediante “feature importance”, de los comportamientos más importantes podemos destacar los siguientes:

1. Que una oportunidad se encuentre dentro del rango de 0 y 50.000 \$ influye cerca de un 7%. Por lo general afecta de manera positiva a “Won”.
2. Que una oportunidad sea del grupo de producto 2875 influye cerca de un 6%. Por lo General afecta de manera positiva a “Won”.
3. Que una oportunidad tenga una duración menor a 1 mes, influye cerca de un 5% Por lo general afecta de manera positiva a “Won”.
4. Que una oportunidad pertenezca a “Substation Automation System” influye cerca de un 3%. Por lo general afecta de manera positiva a “Lost”.
5. Que una oportunidad se encuentre en una duración de entre 4 a 8 meses, influye cerca de un 2%. Por lo general afecta de manera positiva a “Lost”.

Es importante señalar que estos comportamientos se obtienen de “Feature Importance” queriendo decir que estos 5 comportamientos están influyendo en la capacidad del modelo en realizar predicción y ese porcentaje no es necesariamente para todas las oportunidades que se ingresan.

Además, realizar el Feature Importance y el análisis exploratorio de datos nos permitió seleccionar las 12 variables para realizar predicciones, las cuales son:

1. Account Customer
2. Age Range
3. Channel Class Level 1
4. Rango de Precios
5. Solution Line
6. End Usage Level 1
7. Related Product Group
8. Description
9. End Usage Level 2
10. Application
11. Opportunity Owner
12. Year

Con respecto al objetivo general del proyecto, se puede apreciar que se cumplió de buena manera, se logró crear una herramienta que permite apoyar la toma de decisiones, como lo es un modelo predictivo que permite a los vendedores saber que oportunidades tendrán mayor probabilidad de éxito, el uso de este modelo permitió generar un aumento sobre el 10% con respecto a las 3 comparaciones de meses anteriores lo que indica un cumplimiento del objetivo que se planteó en un inicio.

### 13. Conclusiones

En cuanto a los resultados del proyecto, se puede apreciar que, si bien cumple con los objetivos propuestos en un principio, es necesario poder seguir midiendo su desempeño con mayor número de oportunidades y poder probarlo en distintos tipos, ya sea de productos, precios, categorías, etc. Ya que podría estar sesgado y por lo mismo es importante ir actualizándolo con nuevos datos y probando su desempeño a través del tiempo. Aun así, el resultado que entrega el modelo es el esperado y luego de estudio y pruebas de desempeño cumple con los objetivos requeridos por la empresa y se comporta correctamente a pruebas con distintas oportunidades que se cerraron previamente a la creación del modelo.

Finalmente, para concluir nos damos cuenta de que el proceso de toma de decisiones de “Bid no Bid” y “Go no Go” es un proceso que requiere de mucha información, especialmente para el tipo de oportunidades que se presentan en Hitachi Energy, ya que conllevan una gran inversión de tiempo y de recursos, además de que son oportunidades que generan grandes ingresos para la empresa y no son oportunidades que se presentan todos los días. Es por esto que es muy importante tener un proceso que permita tomar decisiones de manera informada y que nos entregue altos grados de confiabilidad en el futuro resultado de la oportunidad. El proyecto de un modelo predictivo dio resultados positivos en cuanto a las predicciones realizadas y cumplió de muy buena manera con los objetivos propuestos en un principio, pero como pudimos apreciar en el estado del arte, la toma de decisiones siempre tiene que ir acompañada de ciertos análisis de riesgos, requerimientos mínimos de confiabilidad, alineación con los objetivos, capacidades de la empresa, etc. Por lo tanto, es muy importante tener en cuenta que una herramienta como lo es el modelo predictivo, es una herramienta que permite apoyar las decisiones, no es una herramienta para cambiar la toma de decisiones ni para eliminar procesos que ya se estaban realizando anteriormente, sino que viene a entregar mayor información y mayor certeza del resultado que tendrá una oportunidad, pero siempre se debe tener en cuenta que el modelo está basado en la probabilidad y siempre existe por más pequeña que sea la probabilidad, que falle. Si bien una toma de decisiones basada en variables subjetivas puede ser desinformada y no muy precisa, al apoyarla con una herramienta, por más que un modelo predictivo tenga una precisión muy alta, siempre existen externalidades que los datos no pueden analizar ni predecir por lo que una herramienta y un proyecto como este no espera remplazar a un vendedor o a un proceso, sino que busca apoyarlo y complementarlo con información basada en los datos.

#### 14. Referencias

Ahmad, I. (1990). Decision-support system for modeling bid/no-bid decision problem. *Journal of construction engineering and management*, 116(4), 595-608.

De Araújo, M. C. B., Alencar, L. H., & De Miranda Mota, C. M. (2021). Classification model for bid/no-bid decision in construction projects. *International Transactions in Operational Research*, 29(2), 1025–1047. <https://doi.org/10.1111/itor.13037>

Marcano Aular, Yelitza Josefina, & Talavera Pereira, Rosalba. (2007). Minería de Datos como soporte a la toma de decisiones empresariales. *Opción*, 23(52), 104-118. Recuperado en 23 de octubre de 2023, de [http://ve.scielo.org/scielo.php?script=sci\\_arttext&pid=S1012-15872007000100008&lng=es&tlng=es](http://ve.scielo.org/scielo.php?script=sci_arttext&pid=S1012-15872007000100008&lng=es&tlng=es).

Pignano, A. S., & Pino, P. J. (2021). La generación de valor mediante el uso de data science en la toma de decisiones comerciales de tiendas por departamento. Trabajo de investigación de bachillerato, Pontificia Universidad Católica del Perú. Repositorio institucional Pontificia Universidad Católica del Perú. <https://tesis.pucp.edu.pe/repositorio/handle/20.500.12404/19468>

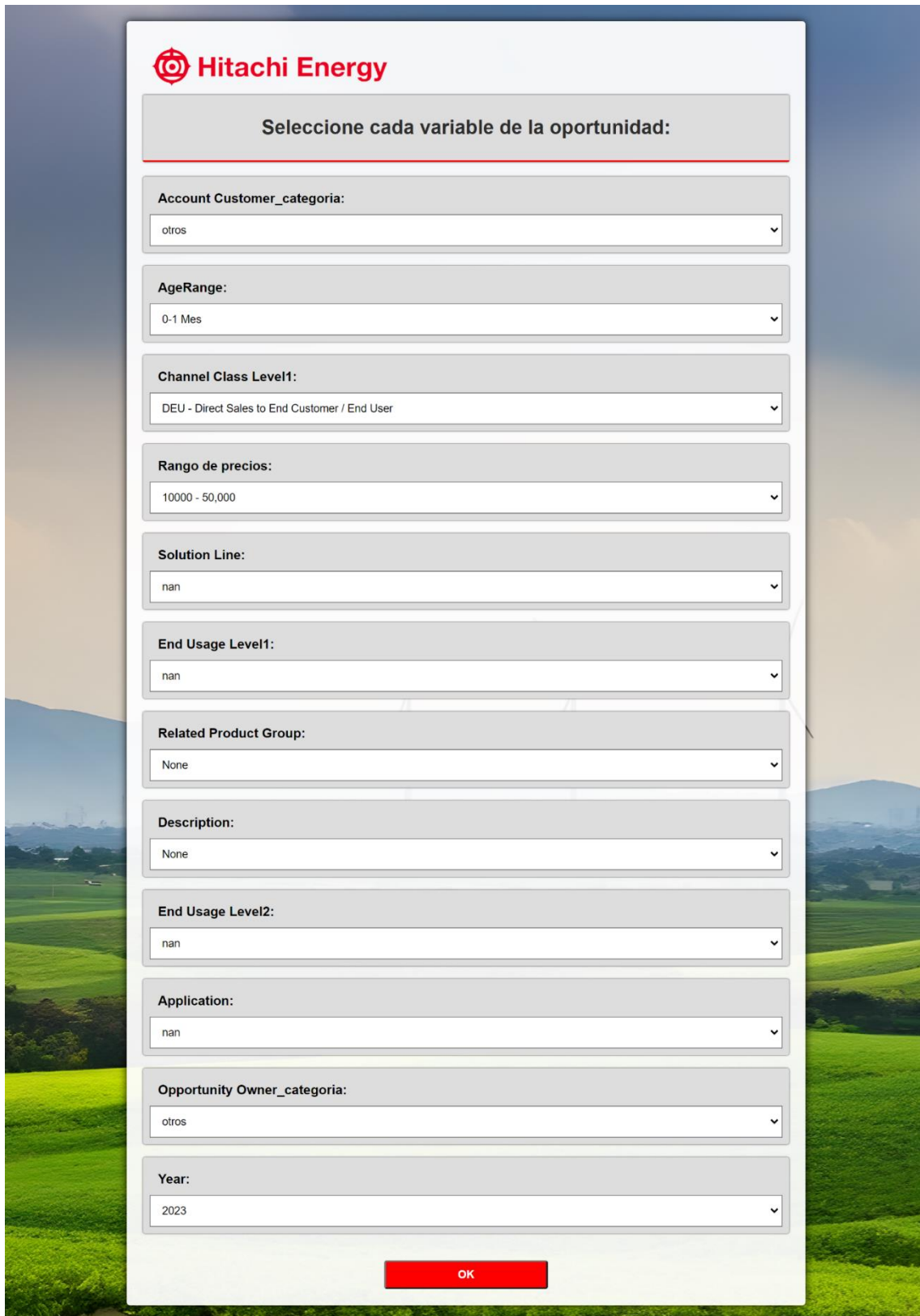
Parra, F. (2019). Estadística y Machine Learning con R. Recuperado de <https://bookdown.org/content/2274/portada.html>

Alaminos-Fernández, A. F. (2022). Árboles de decisión en R con Random Forest. *RUA: Árboles de decisión en R con Random Forest*

Berlanga, V., Rubio Hurtado, M. J., & Vilà Baños, R. (2013). Cómo aplicar árboles de decisión en SPSS. *REIRE. Revista d'Innovació i Recerca en Educació*, 6(1), 65-79. <http://hdl.handle.net/2445/43762>

Galán Cortina, V. (2016). Aplicación de la metodología CRISP-DM a un proyecto de minería de datos en el entorno universitario (Bachelor's thesis).

15. Anexo



The image shows a web form for Hitachi Energy. At the top left is the Hitachi Energy logo. Below it, a header bar contains the text "Seleccione cada variable de la oportunidad:". The form consists of several sections, each with a label and a dropdown menu. The sections are: "Account Customer\_categoria:" with "otros" selected; "AgeRange:" with "0-1 Mes" selected; "Channel Class Level1:" with "DEU - Direct Sales to End Customer / End User" selected; "Rango de precios:" with "10000 - 50,000" selected; "Solution Line:" with "nan" selected; "End Usage Level1:" with "nan" selected; "Related Product Group:" with "None" selected; "Description:" with "None" selected; "End Usage Level2:" with "nan" selected; "Application:" with "nan" selected; "Opportunity Owner\_categoria:" with "otros" selected; and "Year:" with "2023" selected. At the bottom right of the form is a red button labeled "OK".

**Hitachi Energy**

Seleccione cada variable de la oportunidad:

**Account Customer\_categoria:**  
otros

**AgeRange:**  
0-1 Mes

**Channel Class Level1:**  
DEU - Direct Sales to End Customer / End User

**Rango de precios:**  
10000 - 50,000

**Solution Line:**  
nan

**End Usage Level1:**  
nan

**Related Product Group:**  
None

**Description:**  
None

**End Usage Level2:**  
nan

**Application:**  
nan

**Opportunity Owner\_categoria:**  
otros

**Year:**  
2023

OK

Figura 11: Página Web para ingresar una nueva oportunidad



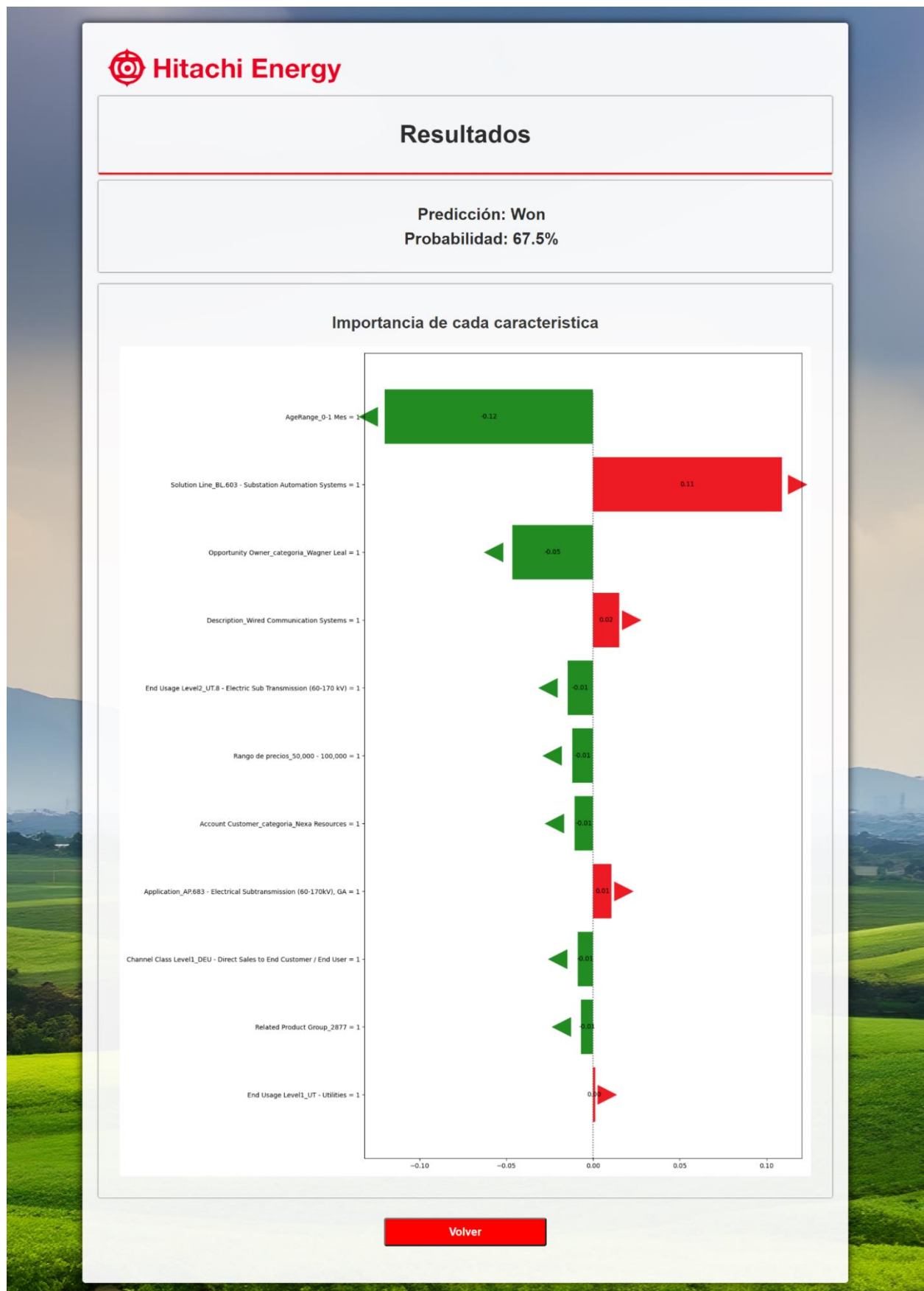


Figura 12: Resultado predicción con importancia de cada característica de la oportunidad ingresada