Generación de Insights para la Identificación de Oportunidades Comerciales para la Venta de Fondos Mutuos Propios

Informe Final

Principal Servicios Corporativos S.A.

Diego Guerra K./20.285.954-2

UNIVERSIDAD ADOLFO IBÁÑEZ

jueves, 7 de diciembre de 2023

1. Resumen Ejecutivo

Principal Chile es una empresa que ofrece servicios financieros, entre sus áreas se encuentra aquella de Distribucion Institucional o DI, cuyo objetivo es la comercialización efectiva de los productos de la empresa a clientes institucionales, cuyas carteras son de alto valor y conforman un porcentaje relevante de los Assets Under Management de la filial.

Dentro de esta área se intenta resolver la pérdida de oportunidades comerciales generadas por procedimientos altamente operativos y anticuados, a través de la automatización del proceso operativo y la implementación de un modelo predictivo que permite clasificar las oportunidades comerciales según sus flujos futuros y verificar los drivers que los impulsaron.

El funcionamiento del modelo es verificado a través de periodos históricos, donde se observa que la capacidad de predecir un 68% de los flujos positivos y un 47% de los flujos negativos relacionados a los fondos, dicho efecto también se evidencia en su implementación, donde se detectan oportunidades de compra que resultan en un incremento de 0,12% de ventas sobre la expectativa del mes, con un tiempo utilizado en generación de insights reducido en un 66%.

Principal Chile is a company that offers financial services, among their divisions is the Institutional Sales division. Their objective is the effective commercialization of the many products and services offered by the company to institutional grade clients, whose portfolios are highly valued and represent a significant percentage of the company's Assets Under Management in Chile.

The objective of the project in this area is to detect missed commercial opportunities caused by antiquated and intensely operative analyses. This is achieved through the implementation of an automation of the repetitive procedures and a predictive model that allows for the classification of the available commercial opportunities by their future flows, and the review of the drivers that drove them.

The model's functionality is verified with historical data, where it is capable to predict a 68% of total positive flows and a 47% of total negative flows related to the funds, this is also corroborated with its implementation, where detected opportunities result in a 0,12% increase in sales over the month's forecast, along with a 66% reduction in time usage.

Índice de Títulos:

Índice de Figuras:

Índice de Tablas:

2. Glosario

- **Machine Learning:** Son metodologías que son capaces de aprender y adaptarse por sí solos para cumplir un cierto objetivo o análisis a partir de modelos estadísticos y matemáticos.
- **AGF, *Administradora General de Fondos:*** Son sociedades anónimas que administran capitales de terceros a través de la oferta de distintos productos como fondos mutuos, fondos de inversión y otros instrumentos financieros.
- **FF.MM., *Fondos Mutuos:*** Son acumulaciones de capital administrados con un cierto objetivo o tesis de inversión, con un objetivo de generar un nivel de rentabilidad con un cierto nivel de riesgo.
- **AUM, *Assets Under Management:*** Corresponde a la cantidad de capital administrado por una AGF en su totalidad o el tamaño de un fondo o inversión.
- **AC, *Asset Class:*** Corresponde a una clasificación que caracteriza a las inversiones según la naturaleza del activo subyacente en el que se invierte, por ejemplo, la inversión en participación de empresas se denomina renta variable, mientras que la adquisición de bonos se denomina renta fija, cada una de estas categorías tiene más subdivisiones que caracterizan al fondo que componen.
- **SS, *Super Sector:*** Corresponde a una caracterización específica de los tipos de inversiones que componen a un fondo, compuesto por la procedencia y por la Asset Class del fondo en cuestión, entonces ejemplos de Super Sectors podría ser bonos de mercados emergentes europeos, acciones estadounidenses, etc.
- **API, *Application Programming Interface:*** Corresponde a una interfaz entre una aplicación local con otra externa, que provee ciertas funcionalidades o datos a través de comandos específicos implementados en la aplicación local.
- **Players:** Corresponde a una terminología utilizada para los partícipes relevantes en el entorno de clientes institucionales, generalmente denotando a instituciones tradicionales, con años de trayectoria y destacados en la industria.
- **Track record:** Consiste en el tiempo de vida que tiene un fondo particular, generalmente los clientes institucionales no consideran fondos que no tienen un track record de al menos tres años.

3. Introducción

Principal Servicios Corporativos o Principal Chile es la presencia en el país de la empresa financiera *Principal Financial Group*, número 266 en el reconocido ranking Fortune 500 estadounidense. Esta es fundada en Des Moines, Iowa en el año 1879 como un proveedor de seguros de vida, hoy se sigue desempeñando en dicho rubro y ha expandido su ofrenda a la gestión de fondos de pensiones y gestión de inversiones generales. Su presencia en Chile está marcada por su rama de seguros de vida, su administradora de fondos de pensiones y su administradora general de fondos, las cuales son supervisadas por el Country Head, cuya labor consiste en establecer la visión estratégica de la casa matriz localmente.

El área donde la pasantía se lleva a cabo se llama Distribución Institucional, es un grupo inherentemente comercial cuyo rol corresponde a la comercialización de los productos y servicios que ofrece la empresa, particularmente a clientes de grado institucional, es decir, empresas con carteras de capitales medidas en miles de millones de pesos chilenos. Estas son altamente atractivas para la empresa, por lo que se ofrece un servicio íntegro, personalizado y de excelencia para captar y retenerlos en el tiempo.

La composición del equipo privilegia la eficiencia operacional y la agilidad por sector de negocio, la estructura de este es la siguiente.

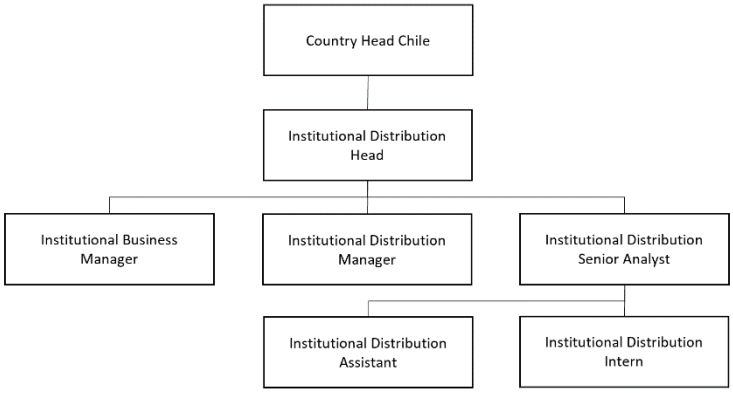


Figura 1. Organigrama de Distribución Institucional en Principal AGF

3.2 Contexto del problema

El problema detectado en el área afecta al proceso de la identificación de oportunidades comerciales en la venta de fondos mutuos propios de Principal AGF dentro del mercado institucional chileno. Para apoyar el proceso de identificación de oportunidades de venta, actualmente se utiliza una herramienta basada en Excel que utiliza reportes mensuales de las carteras de las gestoras de fondos en Chile provenientes de FundPro, una empresa que provee a Principal con detalles de las carteras de clientes institucionales andinos, e información de fondos a través del Add-In de Excel de Morningstar.

Esta herramienta genera un resumen descriptivo de la información contenida en un informe individual de FundPro, permitiendo ver el estado actual de la industria respecto a los flujos y holdings de los clientes institucionales en Chile, e incluyendo una hoja para comparar los rendimientos de los fondos en dichas carteras con aquellos propios de la empresa para detectar oportunidades comerciales manualmente.

Pero ¿Qué se entiende por una oportunidad comercial o de venta? Esto puede definirse como la presencia de un fondo cuyo rendimiento no es óptimo según diversas métricas, en la cartera de una institución en el mercado chileno, es decir, tiene capital invertido ligado a este activo que no está generando el mayor beneficio para la institución y sus propios clientes. En esa situación, si Principal tiene un fondo dentro de la misma categoría (Super Sector) con

rendimiento más atractivo, existe una oportunidad comercial que debe ser identificada y explotada por el equipo de DI.

Entonces para entender de mejor manera el proceso en el cual se involucra la realización del proyecto, se levanta un diagrama de las acciones que llevan a cabo día a día para la identificación y cuantificación de oportunidades comerciales en el mercado institucional chileno.

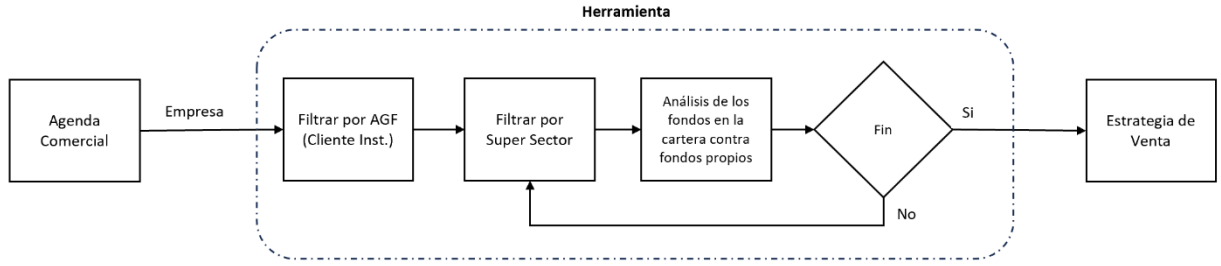


Figura 2. Diagrama de proceso actual

El diagrama anterior contempla el uso simplificado de la herramienta, que depende de una aproximación de las oportunidades comerciales presentes para la confección de la agenda comercial, es decir, se deben establecer reuniones sin confirmar con exactitud la presencia de oportunidades comerciales para realizar la venta. Asimismo, presupone que el cliente a quien se ofrecerá el fondo propio es la AGF considerada en el flujo, lo cual se cumple solo para porcentaje pequeño de los clientes totales considerados. Para el resto se deben considerar varias AGFs en el análisis para verificar las tendencias que marcan al mercado entre el periodo pasado y actual.

En su uso simplificado, la herramienta ya presenta un bucle que genera un flujo de diez a quince minutos por AGF filtrada, con dos a tres de ellos siendo dedicados solamente a refrescar los datos desde Morningstar al alterar los filtros.

Dado que el propósito de este análisis corresponde a realizar un escaneo de los grandes *players* de la industria institucional, la verdadera utilidad de este análisis sería obtener la información consolidada de todas las AGFs incluidas inmediatamente, este flujo con la herramienta actual tardaría alrededor de quince horas en realizar y se vería de la siguiente manera.

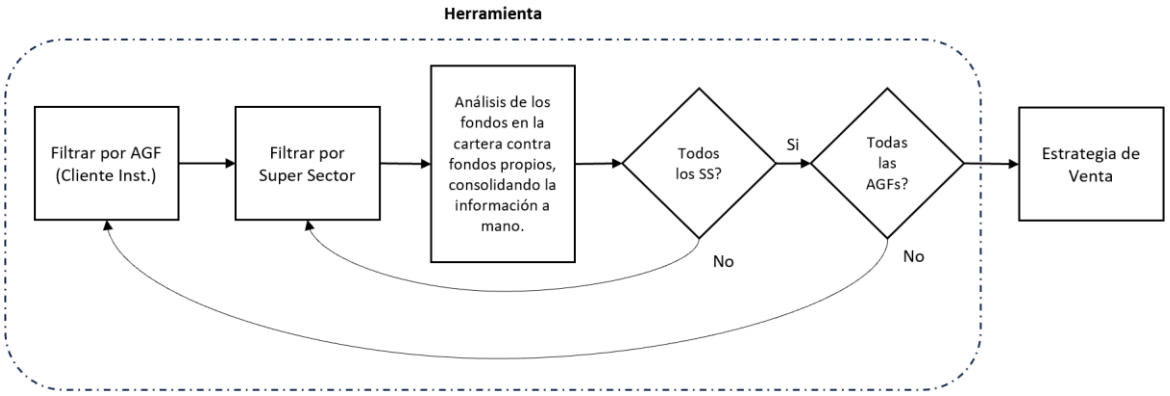


Figura 3. Diagrama de proceso de la herramienta antigua con análisis completo

Sin embargo, el dolor más importante que acontece al área consiste en dolores no cuantificables directamente, como las pérdidas asociadas a la no identificación de oportunidades de negocios en un tramo de tiempo oportuno para accionarla o el compromiso de tiempo y recursos a oportunidades subóptimas.

Entonces, para identificar el problema a partir de los dolores identificados en el área, se utiliza la metodología de los “5 Why’s”, la cual consiste en el cuestionamiento iterativo de la fuente de los problemas para llegar a la fuente de ellos.

Dolor: Existen oportunidades de venta de fondos mutuos de Principal en el mercado institucional chileno perdidas.

- **¿Por que?** Porque no se sabe oportunamente que fondo mutuo venderle a quién.
  - **¿Por que?** Porque los reportes actuales no transmiten dicha información.
    - **¿Por que?** Porque funcionan a partir de filtros manuales y no consideran toda la información útil para ello.
      - **¿Por que?** Porque existen limitaciones asociadas con el ingreso de información, a la herramienta y a al alcance de los reportes, simplificando la calidad del análisis.
        - **¿Por que?** Porque existen aspectos intertemporales relevantes al análisis, que sólo pueden ser detectados y analizados con una herramienta más poderosa.

Figura 4. Diagrama de 5 Why's (5 por qués).

Con la metodología anterior, se puede verificar la presencia de información valiosa no solo en la realización de un análisis de la situación en un momento *t* cualquiera, sino que, en la agregación de estos distintos puntos en el tiempo para lograr extraer un valor mayor del actual, reflejándolo en los KPIs del área.

De esta misma manera, podemos definir formalmente el problema de la siguiente manera: *"El análisis de FF.MM. en carteras de AGFs chilenas realizado por el área institucional de Principal Chile se encuentra limitado por la herramienta utilizada, resultando en un uso lento y no aportando a la identificación de oportunidades comerciales presentes."*

Entonces para detectar el verdadero alcance en el área a partir del problema planteado, se utiliza la metodología del diagrama de Ishikawa, que permitirá listar los aspectos relevantes a considerar en los pasos siguientes del desarrollo del proyecto, ubicado en el anexo 13.1.

A partir del diagrama elaborado, se identifican cinco áreas importantes para el análisis de los factores asociados con el problema, los cuales deben ser considerados al llevar a cabo la solución de este.

4. Objetivos

En la siguiente sección, se encuentra el desglose de los objetivos, tanto principal como específicos para el cumplimiento del propósito del proyecto.

4.1 Objetivo general

El objetivo general que pretende cumplir el proyecto se puede sintetizar en la siguiente frase.
"Generar insights comerciales que permitan al área de distribución institucional de Principal Chile identificar oportunidades de venta de fondos mutuos propios a otras instituciones gestoras de fondos en Chile de una manera rápida y simplificada, permitiéndoles asignar tiempo y recursos eficientemente."

4.2 Objetivos específicos

- Reducir costos asociados a tiempo en un 50% respecto a la situación actual.
- Incrementar las ventas de fondos mutuos propios en un 0,14% mensual o 1,7% anual.

4.3 Medidas de desempeño

- **Métricas asociadas al tiempo:** Permiten medir el grado de optimización del flujo de uso de la herramienta.
    - *Variación en tiempo de procesado y preparación de datos vs. la situación medida inicial* **[%]**
- **Métricas asociadas a la calidad del análisis:** Permiten medir los efectos de los cambios en la capacidad de ventas del ejecutivo.
    - *Flujos generados con apoyo de información de la herramienta vs. las ventas totales del mes* **[%]**

4.4 Planificación

Habiendo definido los objetivos necesarios para llevar a cabo la solución, se puede establecer una planificación que rija los plazos necesarios para cumplir estas actividades y avanzar de manera ordenada y prolija en la solución del problema. Entonces la planificación se puede resumir en la siguiente Carta Gantt.

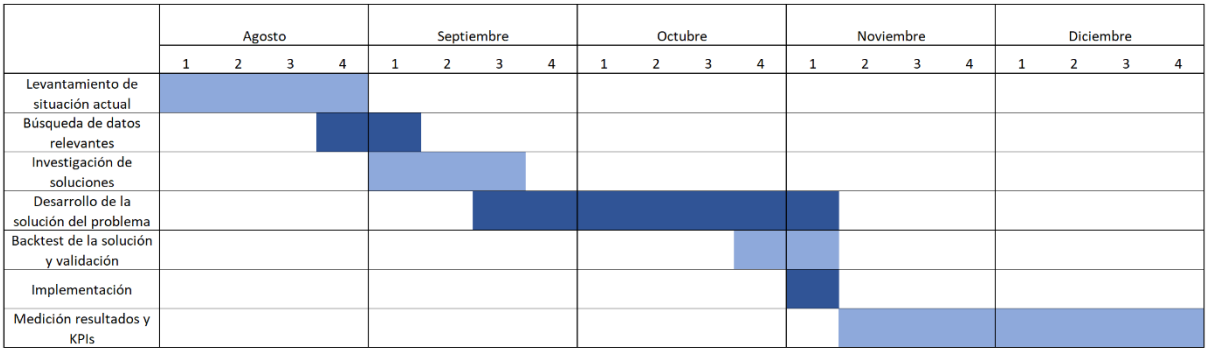


Figura 5. Carta Gantt del proyecto

5. Estado del Arte

Para acotar las posibilidades para resolver el problema previamente descrito, se realiza un análisis del estado del arte que permita comprender como este problema ya ha sido resuelto en el mercado, y las alternativas que existen.

Dado que el problema consiste en la inhabilidad de detectar la totalidad de las oportunidades comerciales existentes en el mercado, se proponen las siguientes soluciones tipo que deben cumplir con los siguientes requisitos

Sin embargo, el producto que se pretende utilizar debe ser altamente personalizable para el propósito que se quiere lograr en respecto a este uso, por ende, se analizan como opciones del mercado a plataformas de análisis de datos existentes en el mercado en virtud de sus ventajas y desventajas.

5.1. Avances Académicos Relevantes

Como metodologías utilizadas hoy en día en la academia que podrían ayudar en la solución del problema que acontece al área, se consideran tres papers que tratan sobre la predicción o estimación de factores relacionados a ventas a través de metodologías de *data science* y *machine learning.*

1. ***Regression as classification***[1]*: Explica el planteamiento de un problema de predicción de nivel como un problema de clasificación multiclass, donde la codificación ofrece mejores resultados que su contraparte regresiva.*
2. ***Comparison study using ARIMAX and VARX in cash flow forecasting***[2]*: Propone la utilización de modelos de series de tiempo para pronosticar los flujos y niveles de dinero físico para empresas bancarias a partir de la contingencia generada por el Covid-19.*
3. ***A Study of Demand and Sales Forecasting Model Using Machine Learning***[3]*: Propone el planteamiento de algoritmos de machine learning para la clasificación de reviews según el sentimiento percibido en mercados de retail.*

Para efectos de la solución del problema del área, se opta por adoptar características del primer y tercer artículo considerado, optando por convertir los flujos observados en los informes de FundPro en variables categóricas representativas si un flujo es positivo o negativo, para así realizar predicciones más acertadas y reconocer los patrones que definen estos movimientos.

Debido a que lo descrito anteriormente consiste en una metodología y uso específico, no existen herramientas presentes en el mercado actual que se adecuen para la realización de este procedimiento, por lo que este tendrá que ser desarrollado e implementado en plataformas de

[1] R. Salman and V. Kecman, "Regression as classification," *2012 Proceedings of IEEE Southeastcon*, Orlando, FL, USA, 2012, pp. 1-6, doi: 10.1109/SECon.2012.6196887.
[2] Andreas, Christopher & Faricha, Anifatul & Ulyah, Siti & Susanti, Rika & Mardhiana, Hawwin & Achirul Nanda, Muhammad & R., Firman. (2022). Comparison study using ARIMAX and VARX in cash flow forecasting. AIP Conference Proceedings. 2641. 030023. 10.1063/5.0118519.
[3] Aneesh Tony, Pradeep Kumar, Rohith Jefferson, Subramanian, "A Study of Demand and Sales Forecasting Model Using Machine Learning", Psychology and Education, 2021.

data science como Python o R. Para ello es que a continuación se considerarán las plataformas relevantes que permitan la extracción de datos financieros a utilizar en el modelo desarrollado y posean también un servicio de hosting de código o API compatible con Python.

5.2. Análisis de Entornos de Programación

Para la implementación de la solución, se opta por desarrollarse en un entorno cloud ofrecido por participantes del mercado de datos financieros, ya que permiten acceder a su información de manera nativa, no requieren de instalación de librerías o software adicional y permiten almacenar otras fuentes de datos para ser utilizados en los análisis creados, las herramientas en el mercado actual relevantes a este proyecto son las siguientes.

Aplicación	Instalación	Documentación	Soporte	Compatibilidad	Requisitos	Puntaje
Analytics Lab (Morningstar)	5	5	5	5	5	5
BQuant (Bloomberg)	5	5	5	5	3	4,6
Developer's Toolkit (Factset)	1	3	4	1	2	2,2
API Genérica (Yahoo Finance)	3	3	1	1	1	1,8

Tabla 1. Evaluación de entornos de programación de proveedores de datos

5.3. Análisis de Proveedores de Datos Financieros

Por otra parte, se debe realizar un análisis a la fuente de datos que se utilizará para la solución, con el fin de proponer el producto que se adapte de mejor manera a las limitaciones de las plataformas y a las necesidades del área, entre otras.

Aplicación	Tamaño y calidad de datasets	Compatibilidad	Soporte	Costo	Requisitos	Puntaje
Bloomberg	5	4	3	2	3	3,4
Morningstar	4	5	4	3	5	4,2
Factset	3	3	2	3	3	2,8
Yahoo Finance	2	4	3	5	1	3,0

Tabla 2. Evaluación de proveedores de datos

5.4. Conclusión de Análisis de Plataformas

A partir de las tablas anteriores, se puede obtener la siguiente tabla resumen donde se le suman los puntajes asociados a las plataformas y fuentes de datos financieros.

	Morningstar	Bloomberg	Factset	Yahoo Finance
Puntaje total	9,2	8,0	5,0	4,8

Tabla 3. Resumen de herramientas y proveedores de datos

A partir de ello es que se opta por realizar el análisis en la plataforma de Analytics Lab de Morningstar, ya que esta permite almacenar los datos en su misma nube, importar datos nuevos y trabajarlos en un entorno con extensa documentación. Particularmente, el área de DI ya posee una licencia para este software, lo que se incluyó esta característica en los puntajes por medio de la columna de requisitos.

6. Riesgos Asociados al Proyecto

Para hacer un análisis íntegro de los efectos de la implementación de esta solución en el área, se realiza un desglose de posibles riesgos, con un fin de implementar las mitigaciones asociadas en conjunto con la solución.

Riesgo	Probabilidad de ocurrencia	Mitigación
Incapacidad de arreglo o modificación de la herramienta.	Alto	Entregar en conjunto a la herramienta, documentación y código comentado que permita entender el funcionamiento línea por línea para entender la herramienta, utilizable por el área o por los grupos de TI, BI o Estudios.
Cambio de formato de los datos de reportes de carteras nuevos.	Medio	A pesar de que ya hubo un cambio en enero de 2023, la herramienta utiliza solo columnas que se han mantenido constantes para los periodos estudiados.
Resistencia a la adopción de la herramienta.	Bajo	El área ya utiliza rutinariamente Morningstar, por lo que todo uso adicional de la plataforma provee mayor valor a la empresa por su gasto recurrente.
Predicciones realizadas con un modelo con bajo ajuste.	Medio	Para mitigar los riesgos asociados a las predicciones con un fit no satisfactorio, se utilizan los datos de 12 meses históricos de transacciones de mercado, con datos de rentabilidades a 3 años para maximizar el ajuste. Asimismo, el modelo produce un reporte con los resultados de la clasificación de prueba y las probabilidades asociadas a las etiquetas, para que la decisión final recaiga en el criterio del usuario de la herramienta.
Cambio de proveedor de información de fondos	Media-baja	En caso del cambio de proveedor de información, se conoce la fuente de donde FundPro obtiene los datos desde la CMF de los distintos fondos, sin embargo, no es labor menor como para abandonar su servicio actualmente.

Cancelación del servicio de Morningstar	Bajo	El área de DI no es la única que utiliza el servicio, este es utilizado por Estrategia de inversiones y desarrollo de productos, y es un servicio estándar en la industria. Sin embargo, de ser cancelado, se podrá migrar a BQuant con relativa facilidad.

Tabla 4. Identificación y mitigación de riesgos

7. Evaluación Económica

En la siguiente tabla, se muestran los costos asociados a tiempo que conlleva el uso de la herramienta en una base mensual y anual.

	Preparación e inserción de datos	Generación de Insights	Mantención y arreglo
Tiempo	30 min/mes	3,75 hrs/semana	40 hrs/mes[4]
Tiempo Total (Mes)	15,5 hrs		
Tiempo Total (Año)	186 hrs		

Tabla 5. Tiempos asociados al uso de la herramienta

Para determinar los efectos económicos de la implementación del proyecto en la empresa, se lleva a cabo la metodología de flujos descontados en el tiempo. Esta consiste en la actualización o "traer al presente", flujos de caja que serán percibidos en el futuro. Esto se realiza mediante una tasa de descuento que permite identificar cual es el valor actual de estos flujos futuros, la

[4] Corresponde a la situación para el semestre 2023/01, donde la herramienta deja de funcionar y contempla el tiempo utilizado para arreglarla.

noción detrás de esta metodología es la siguiente: "Si el valor presente de el flujo en N años PV crece a una cierta tasa de descuento α, resulta en el flujo de caja futuro FV". Lo mismo puede ser expresado con la siguiente fórmula.

$$FV = PV * (1 + \alpha)^N \quad \Longrightarrow \quad PV = \frac{FV}{(1 + \alpha)^N}$$

De esta manera, para cada año se emplaza el valor presente del flujo base considerado al año 1, descontado por la tasa de descuento α = 15% por la cantidad de años que corresponda, donde la sumatoria de ello resulta en el NPV o valor presente neto del proyecto. Por otra parte, para contabilizar los gastos de tiempos se presupone un sueldo mensual de MM CLP$ 2,5, sin embargo, este no es representativo de la situación real y es utilizado para demostrar los efectos monetarios pronosticados del proyecto.

	Variación Esperada (%)	**Año 0**	**Año 1**	**Año 2**	**Año 3**	**Año 4**	**Año 5**
		(CLP $)					
Tiempo ahorrado	(50%)	-	937.500	796.875	677.344	575.742	489.381
Incremento en ventas	1,70%	-	1,70%	1,45%	1,23%	1,04%	0,89%
Costo de desarrollo	-	(400.000)	-	-	-	-	-
Total por periodo ($)		(400.000)	937.500	795.875	677.344	575.742	489.381
Ventas Adicionales Descontadas		***6,46%***					
Valor presente del proyecto por flujos fijos.		***3.492.816***					

Tabla 6. Evaluación económica del proyecto a 5 años

El ejercicio anterior resulta en un factor variable basado en el nivel de ventas actuales de un 6,46% a lo largo de los próximos años a valores actuales, con un orden de magnitud de cientos

de miles de dólares. Por otra parte, los tiempos ahorrados generan un ahorro fijo de aproximadamente CLP$ MM 3,5, sin considerar que dicho tiempo ahorrado puede ser invertido en la generación de mayor rentabilidad por otras vías u otras actividades beneficiosas para el negocio.

8. Metodología

8.1. Objetivo

El objetivo de la implementación del modelo de clasificación corresponde en la predicción de los flujos futuros a través de su codificación a variable categórica, permitiendo obtener mejores resultados en la predicción final.

8.2. Desarrollo CRISP-DM

Entonces para llevar a cabo el desarrollo del proyecto, se utilizará la metodología de CRISP-DM (Cross Industry Standard Process for Data Mining). Esta metodología se divide en seis etapas a considerar para la implementación de un modelo de data science, las cuales son expandidas a continuación.

8.2.1. Etapa 1: Business Understanding

Esta etapa corresponde a generar el entendimiento de qué es lo que requiere el negocio como resultado de este proceso, se definen objetivos y una planificación para la realización de este, lo cual ya ha sido abordado en las secciones anteriores de este informe.

8.2.2. Etapa 2: Data Understanding

Corresponde a la colección e interiorización con los datos a utilizar para cumplir con los objetivos establecidos en la etapa anterior, identificando los datos disponibles, describiéndolos y detectando el estado de ellos para así definir los pasos a tomar para limpiarlos y procesarlos.

8.2.3. Etapa 3: Data Preparation

Corresponde a la etapa más importante del proyecto y el volumen más grande de este, corresponde a la selección, limpieza, estructuración, integración y formateo de los datos disponibles para consolidarlos en un set de datos a utilizar en el modelamiento.

8.2.4. Etapa 4: Modeling

Esta etapa corresponde a el uso del set de datos generado para modelar una variable dependiente establecida, considerando la selección de los modelos a considerar, la creación de los modelos y su posterior comparación para realizar la selección final de este.

8.2.5. Etapa 5: Evaluation

Esta etapa corresponde a la evaluación del cumplimiento de los objetivos de negocio establecidos inicialmente utilizando el modelo desarrollado, donde se determina si continuar con la producción o iterar nuevamente el proceso.

8.2.6. Etapa 6: Deployment

En caso de haber seleccionado el modelo y cumplido los objetivos planteados inicialmente, se procede a distribuir los resultados del proceso desarrollado, sea un reporte o un pipeline complejo.

9. Implementación

En esta sección, se desarrollará el proceso de entendimiento y preparación de los datos disponibles, como la evaluación de modelos correspondientes a las etapas 2, 3 y 4 de la metodología CRISP-DM, concluyendo con la introducción de la medición de la etapa 5 de la evaluación de los resultados.

9.1. Librerías Utilizadas

Para la creación de la herramienta, se utilizan algunas librerías implementadas en la plataforma de Morningstar Analytics Lab, sin embargo, la herramienta se desarrolla primariamente con código propio y contempla aproximadamente 450 líneas para su ejecución.

A continuación, se enumeran las librerías utilizadas para las transformaciones de datos y el entrenamiento de modelos, con sus respectivas funcionalidades.

- Numpy / Pandas
- Scikit-learn
- Morningstar Data

9.2. Fuentes de Datos

Las fuentes de datos utilizadas, sean diccionarios o información externa, se encuentran enumeradas a continuación.

- Informes históricos de FundPro, extraídos en ventanas móviles de 13 meses según la fecha de inicio configurada por el usuario, los cuales contienen los flujos por mes de actores institucionales desglosados por fondo y super sector con sus respectivos identificadores.
- Diccionario de SecId, que permite compatibilizar entre los identificadores presentes en los informes de FundPro y aquellos utilizados por Morningstar.
- Retornos históricos mensuales obtenidos según los identificadores observados en los informes de FundPro a través de Morningstar.
- Base con variables exógenas relevantes para el mercado por valores mensuales desde la base de datos estadística del Banco Central de Chile.

9.3. Funciones Utilizadas

Con el incentivo de crear una herramienta eficiente en uso de los recursos disponibles, se opta por la creación de dos funciones principales que son relevantes para remover el efecto

temporal presente en los reportes FundPro y tratar a las observaciones de la misma manera en el modelo.

Para entrenar el modelo con suficientes muestras, es que se utiliza la variable "lag", que pretende determinar a cuantos periodos en el pasado corresponde dicha observación de flujos. Lo anterior para tomar las ventanas de retornos de tres años correspondiente y la información de mercado exógena al momento de que los clientes toman la decisión de compra o venta del respectivo fondo, con el propósito de eliminar el efecto temporal ligado a los datos y poder tomarlos como observaciones independientes.

Las funciones utilizadas para lograr lo descrito son las siguientes.

Nombre	CalcMetrics
Inputs	Index, base de rendimientos y lag asociado a la observación.
Output	Retornos y volatilidades según el lag provisto.
Descripción	El objetivo de la función consiste en la agrupación de los retornos mensuales presentes en la base de retornos en tramos de seis meses, calculando adecuadamente los rendimientos acumulados y las desviaciones estándares de ellos, esto según el lag que define el punto de inicio para estos tramos, considerando que los retornos mensuales se encuentran en columnas, con la más reciente al final.

Tabla 7. Descripción de función CalcMetrics

Nombre	CalcExogs
Inputs	Base de exógenos
Output	Variaciones 1M y 1Y de cada variable exógena en la base.
Descripción	El objetivo de esta función consiste en calcular las variaciones porcentuales de cada regresor exógeno presente en la base y para cada lag.

Tabla 8. Descripción de función CalcExogs

9.4. Preparación de los Datos

Para la implementación de la herramienta, el primer paso corresponde en la consolidación de los archivos de flujos de FundPro en uno solo, considerando su temporalidad. Para ello se utilizan "lags" que indican a cuantos periodos atrás relativo al presente pertenece cada observación en el dataset. Para el periodo actual según fecha cronológica, se establece el lag 0 y se incrementa según corresponda.

A partir de estos lags, es que se calculan las columnas de "Flujo t+1", generando subsets del dataset a partir de que las observaciones cumplan con las siguientes características que aseguran que la observación encontrada sea única y coincidente con lo que se busca.

- La AGF sea la misma que en la observación.
- El fondo sea el mismo que en la observación, a partir de su identificador.
- El mes sea el anterior que en la observación.
- El año sea el mismo si el mes previo a aquél de la observación no genera un cambio de año, o el año anterior e.o.c.

Teniendo los flujos de los fondos en su periodo siguiente, se procede a codificarlos según el siguiente detalle para generar la columna de señales "Señal t+1", la cual será la variable dependiente categórica para realizar la predicción de clasificación con el modelo.

Flujo t+1 (Numèrica)	**Señal t+1 (Categórica)**
Positivo	Flujo Positivo
Negativo	Flujo Negativo
Cero	Flujo Neutro
NA	NA

Tabla 9. Codificación utilizada para flujos futuros

El siguiente paso consiste en añadir la información exógena procesada con la función CalcExogs a set de datos a través de un merge que utiliza los lags como llave para realizar la

equivalencia, insertando la información de mercado al mismo tiempo en que se realiza la decisión de compra o venta.

Posteriormente, se inserta para cada fondo la información de su rendimiento a través de la función CalcMetrics, utilizando su SecId, el índice y el lag de la observación, permitiendo añadir las columnas de retornos y volatilidades por semestre a la base.

El procedimiento anterior es realizado para los fondos presentes en los informes y para los fondos propios insertados en tiempo presente, para posteriormente ser considerados en las predicciones realizadas por los modelos entrenados con los datos de transacciones históricas.

9.5. Análisis Descriptivo de los Datos Finales

Los datos utilizados para entrenar el modelo consisten en solo variables numéricas representativas de variaciones porcentuales, mientras que fuera del entrenamiento de éstos se utilizan variables categóricas como el SecId, los lags y el super sector asociado a cada observación, utilizados para generar el set de datos y entrenar modelos distintos para cada super sector.

El siguiente gráfico muestra el análisis descriptivo del dataset de retornos compuesto anteriormente.

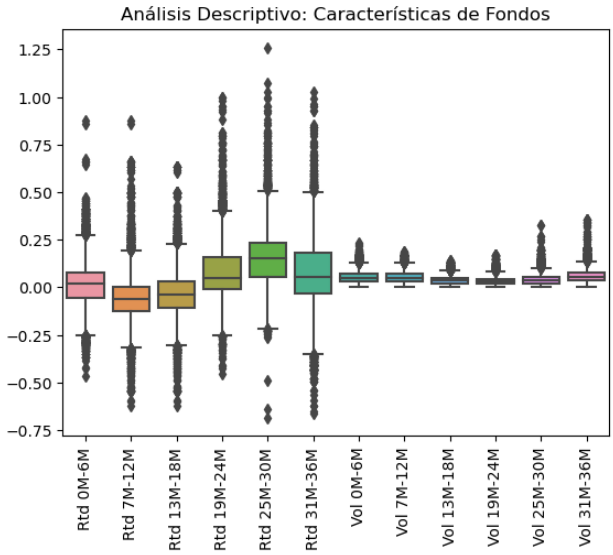


Figura 6. Análisis descriptivo de variables características de fondos mutuos considerados

Por otra parte, se realiza un procedimiento similar con las variables exógenas utilizadas, ya que también deben estar relacionadas al tiempo de apariencia en los informes y deben ser transformadas al formato seleccionado, que en este caso son las variaciones mes a mes, 1M o MoM y variaciones año a año, 1Y o YoY. Dado que las variables exógenas solo dependen del tiempo asociado, es que se calculan todos los lags necesarios inmediatamente, sin necesitar calcular fila por fila.

El siguiente gráfico muestra el análisis descriptivo del dataset de variables exógenas descrito anteriormente.

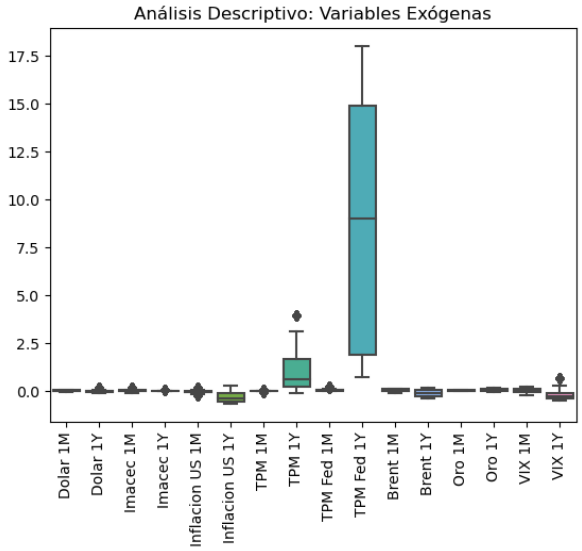


Figura 7. Análisis descrptivo de variables exógenas consideradas

Es necesario destacar que las anomalías observadas tanto para la TPM chilena 1Y como para la TPM US 1Y no son errores, sino que para el periodo estudiado la TPM en chile ha incrementado en un 100% en promedio, mientras que en EE.UU. ha incrementado en casi un 1000%, coincidiendo con la realidad que se pretende reflejar en el modelo. Los efectos de esto se verán reflejados en la magnitud de los coeficientes asociados a dichos regresores, pero no afectará el fit general del modelo dado que los cambios en las políticas monetarias de ambos países son graduales y altamente representativos del sentimiento del mercado.

Dado que los flujos de compra y venta no están necesariamente balanceados, es que el modelo se entrena utilizando la opción “balanced”, que permite balancear la variable dependiente según

su frecuencia de aparición. Por otra parte, los valores NA presentes en los datos ocurren únicamente en el dataset de los retornos y son indicativos de que el fondo en cuestión no posee suficiente *track record*, no siendo considerado para el entrenamiento o predicción de la etiqueta de futuros positivos o negativos.

9.5. Modelos a Considerar

Teniendo en cuenta la codificación de los flujos en variable categórica, se seleccionan los siguientes modelos clasificadores para seleccionar aquél que provea los mejores resultados según las necesidades de la empresa e identifique de mejor manera los patrones presentes en los datos.

- Regresión Logística
- SGD Classifier
- Perceptron
- Passive Aggressive Classifier
- SVC
- Linear SVC
- KNN Clasificador (n = 10)
- Extra Tree Classifier
- Extra Trees Classifier
- Random Forest

9.6. Selección del Modelo

Para la selección del modelo a utilizar, se consideran tanto modelos de clasificación multinomiales con los resultados posibles {“Flujo Positivo”, “Flujo Neutro”, “Flujo Negativo”} como modelos de clasificación binarios con resultados posibles {“Flujo Positivo”, “Flujo Negativo”}, para comparar sus rendimientos y evaluar correctamente las ventajas y desventajas de clasificación multinomial contra la binaria.

Entonces se realizan backtests para cada metodología, obteniendo resultados consolidados al aplicar cross-validations que permiten obtener nuevas divisiones de conjuntos de entrenamiento y prueba, resultando en las tablas resumen en los anexos 13.5 y 13.6 respectivamente.

Para evaluar el rendimiento de los modelos se utiliza la métrica de recall de los flujos positivos, dado que se busca privilegiar los resultados del modelo como una primera aproximación de flujos de los fondos, considerando un análisis posterior de los fondos con resultados positivos. Esto significa que analizar fondos de más se convierte en una diferencia marginal de tiempo, opuesto a perder posibles oportunidades comerciales por la no detección de éstas. En otras palabras, el costo de no detectar una posible oportunidad comercial es alto y aquél de la sobre detección es relativamente bajo.

A continuación, se muestran los desempeños en el recall de flujos positivos de los clasificadores multinomiales y binarios, ordenados según sus medias de mayor a menor.

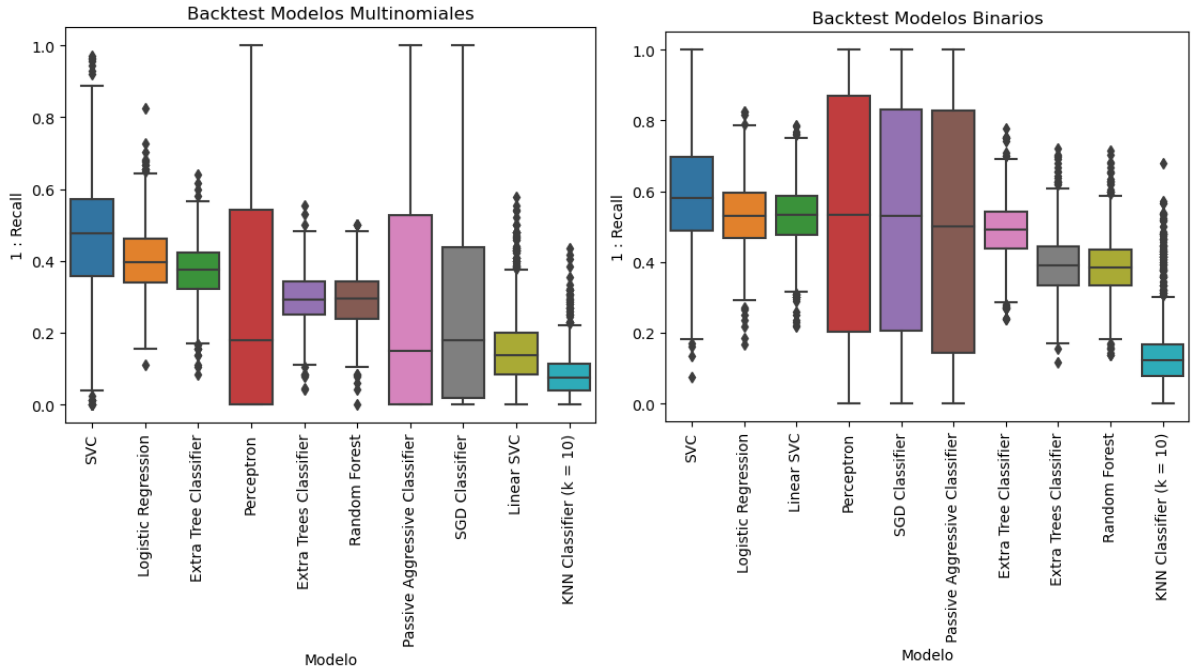


Figura 8. Gráficos de caja y bigotes de los resultados de los backtests realizados, con resultados multinomiales a la izquierda y resultados binarios a la derecha.

A partir de los resúmenes anteriores, se puede notar un desempeño mejor para los clasificadores binarios por sobre los multinomiales, este resultado puede explicarse gráficamente de la siguiente manera.

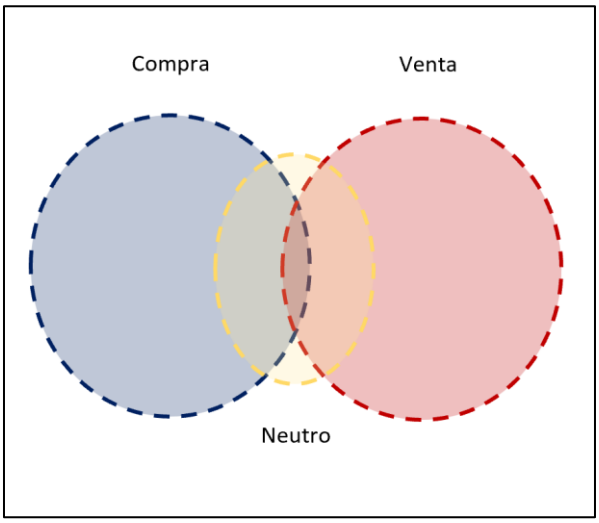


Figura 9. Diagrama de Venn ejemplificando el caso multinomial

Como se puede visualizar, el problema planteado de manera multinomial pretende clasificar a los fondos en los tres casos posibles, la venta, compra y neutro o hold. Esto es codificado como "Flujo Negativo", "Flujo Positivo" y "Flujo Neutro" respectivamente. El problema con esta metodología surge al intentar clasificar como hold o "Flujo Neutro", ya que un fondo con rendimientos por sobre la media pueden ser vendidos o viceversa. Esto ocurre porque los mercados son operados por personas con asimetrías de información, con estrategias individuales para cada participante y con prioridades distintas, todos los anteriores son factores que el modelo no puede ajustar correctamente, castigando a los patrones presentes en los datos y perjudicando el fit de las compras y ventas.

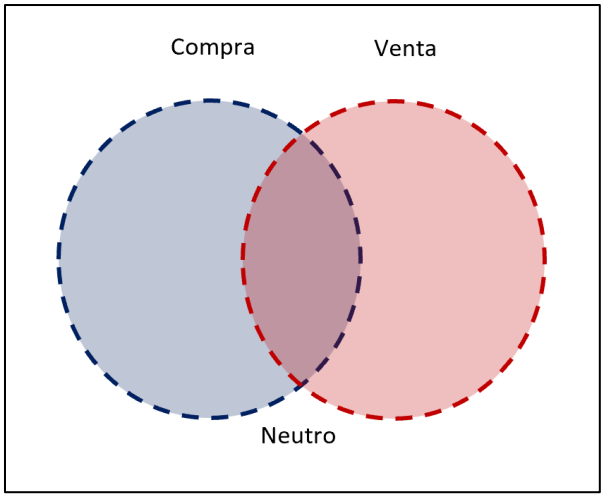


Figura 10. Diagrama de Venn ejemplificando el caso binario

Por otra parte, el modelo binario es capaz de predecir con mayor certeza la compra o venta de los fondos del mercado institucional chileno, como se puede intuir gráficamente y a través de los resultados obtenidos por el backtest. Asimismo, es capaz de generar matices al predecir las probabilidades de cada clase para cada fondo, permitiendo inferir si las características del fondo lo sitúan en los extremos o cerca de la neutralidad, generando así una especie de señal neutra.

Concluyendo la explicación anterior, se muestran en la siguiente tabla los cinco mejores modelos utilizando clasificación binaria, ordenados de mejor a peor recall de flujos positivos.

Modelo	**Medias** **Flujos Positivos: Recall**
SVC	59,6%
Logistic Regression	53,1%
Linear SVC	53,1%
Perceptron	52,7%
SGD Classifier	51,8%

Tabla 10. Tabla resumen de recalls de flujos positivos en backtest binario

A pesar de que SVC haya obtenido el mejor rendimiento en el backtest, se opta por la regresión logística para la clasificación, ya que la pérdida de rendimiento en general se ve compensada por la simplicidad del modelo y su capacidad de generar insights adicionales a partir del análisis de los coeficientes asociados a ella.

Entonces, el procedimiento de uso de la herramienta nueva podría diagramarse de la siguiente manera.

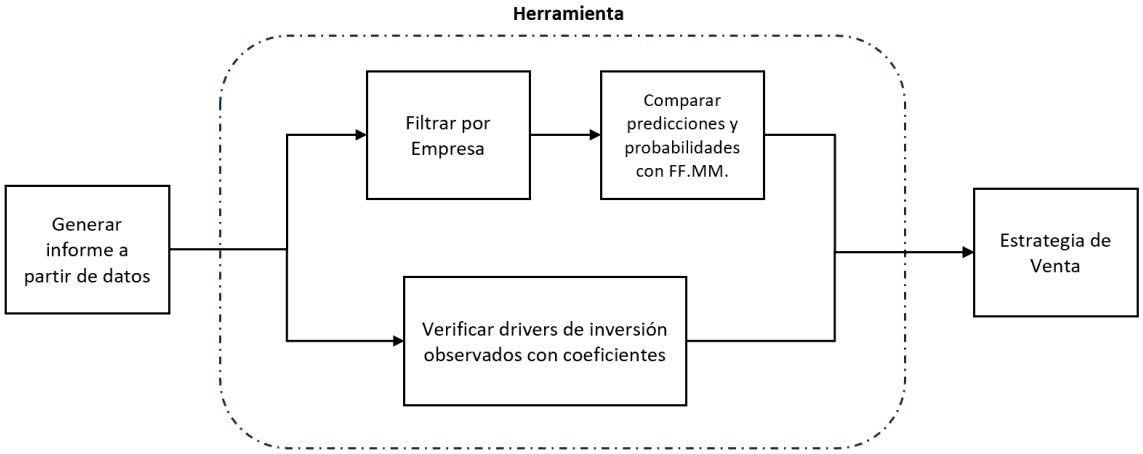


Figura 11. Diagrama de proceso implementando la herramienta nueva

A partir del diagrama de proceso anterior, se puede denotar que no solo se logra eliminar el bucle que estaba presente en la realización de la herramienta, sino que la nueva herramienta permite realizar la actualización de datos y cálculos necesarios para todas las AGFs presentes en los informes de manera consolidada, no requiriendo calcular nuevamente cifras estáticas y obteniendo un overview de las tendencias de los *players* más relevantes del país, ofreciendo *insights* para aquellos incluidos en el análisis como para otras instituciones del mercado.

9.7. Validación de Supuestos

Para validar la selección de las variables exógenas, se grafica la matriz de correlaciones para satisfacer el común supuesto de la falta de dependencia lineal entre regresores, presente en el anexo 13.7.

A partir de las correlaciones obtenidas, se puede concluir que los regresores satisfacen de mejor manera las correlaciones entre los sets de datos probados durante la pasantía, sin embargo, se destaca la presencia de algunas correlaciones relativamente altas, sin embargo, se debe destacar que cada una de las series de datos presentes representa un indicador económico o financiero diferente e inherentemente independiente de los otros.

9.8. Validación del Cumplimiento de Objetivos

Respecto a los objetivos específicos, se midieron los tiempos necesarios para la realización del análisis anterior, tanto para la extracción y preparación de datos como para el tiempo de utilización de la aplicación, considerando los tiempos muertos generados por la actualización de los datos con el Add-In de Morningstar en Excel, tabla ubicada en el anexo 13.3.

La tabla referenciada muestra los tiempos considerados para la ejecución del análisis, donde la extracción y preparación de datos considera la ejecución sin errores de un macro de Excel sobre el reporte descargado manualmente de FundPro, para después ser insertado al reporte a través de Power Query, sin embargo, errores en el reporte o en el macro pueden resultar en la demora más allá de los tiempos medidos.

Por otra parte, el análisis de datos y la cantidad de refrescos se consideran en conjunto, ya que el filtro por *Super Sector* realizado requiere consultar la información de los fondos desplegados a Morningstar para evitar sobrecargar la API, esta extracción se realiza ocho veces en promedio para cada AGF, tomando entonces ocho minutos en promedio que Excel se encuentra inutilizable por la extracción de la información.

Por otra parte, se levantan los flujos observados de los fondos mutuos propios para verificar el cumplimiento de los objetivos asociados a la calidad del análisis y el efecto del proyecto en éstos. Por motivos de confidencialidad, los datos fueron estandarizados y posteriormente graficados en el anexo 13.4.

En el gráfico se muestran las variaciones netas en azul, variaciones positivas en verde y variaciones negativas en rojo, con ganancias acumuladas en amarillo y una estimación lineal de las variaciones netas en la flecha roja. Visualmente, se pueden evidenciar grandes caídas en marzo y octubre relacionadas primariamente a efectos macroeconómicos que afectan al sector financiero, particularmente las guerras que han azotado al mundo últimamente, mientras que en julio se pudieron recuperar una parte substancial de los flujos debido a la calidad de los fondos ofrecidos por la empresa, considerando las condiciones económicas y políticas del momento.

Asimismo, se puede evidenciar una tendencia negativa por los últimos tres meses del año, por lo que se debe considerar la posibilidad que la funcionalidad del modelo no pueda ser comprobada en la dimensión de la variación de las ventas del área.

10. Resultados

La herramienta terminó su etapa de desarrollo a finales de la semana del 30 de octubre de 2023 y entró en ejecución para la semana del 6 de noviembre, generando los siguientes resultados para los flujos de noviembre basado en la información de octubre, vale destacar que, por razones de confidencialidad, los nombres de los fondos no son incluidos en el informe.

En esta sección se desarrollará la conclusión de la etapa 5 y la etapa 6 del modelo CRISP-DM, considerando la evaluación de los efectos del análisis generado en el negocio y la distribución de la herramienta creada a los stakeholders.

10.1. Tiempos Medidos

En un primer lugar, se obtienen las siguientes métricas de tiempo de uso de la herramienta para la medición del objetivo relacionado con optimización de los tiempos.

	Herramienta antigua	**Análisis nuevo**	**Variación porcentual**
Extracción y preparación de datos.	5 minutos	5 minutos	0%
Análisis de datos	15 horas	7,5 horas	50%
Tiempo en actualización de datos por sesión	2 horas, 40 segundos	0 minutos	100%
		Variación total	**66,6%**

Tabla 11. Tiempos ahorrados con herramienta implementada

Las dimensiones anteriores corresponden a las mediciones de tiempos en las siguientes categorías.

- **Extracción y preparación de datos:** Este eje corresponde al tiempo necesario para obtener los datos desde sus respectivas fuentes e ingresarlos a la herramienta. Debido a que se agregan variables exógenas al análisis que deben ser obtenidas desde una fuente no automatizable, es que los tiempos entre situaciones se mantienen constantes.

- **Análisis de datos:** Este eje corresponde al tiempo utilizado por el usuario en analizar los resultados de las herramientas respectivas, debido a que se logra suprimir el tiempo de cálculo de Excel y las repetidas extracciones realizadas con el Add-In de Morningstar, es que el tiempo de análisis de datos se acota a la mitad.

- **Tiempo por actualización de datos:** Este eje corresponde al tiempo que requiere el Add-In de Morningstar en actualizar los datos de los fondos según filtros para acotar la cantidad de tiempo actualizando, la cual tiene un crecimiento exponencial al

incrementar la cantidad de datos solicitada. Debido a que se realiza esta solicitud a través de la API de Morningstar en Python, este tiempo de actualización variable se transforma en una constante considerada en la ejecución de la herramienta, disminuyendo esta métrica en un 100%

10.2. Predicciones Realizadas

Por otra parte, los datos disponibles permiten la generación de las siguientes predicciones en respecto a los flujos futuros de los fondos propios.

Fondo	**Predicción**	**Flujos Netos 11-2023**
Fondo A	Flujo Negativo	Flujo Negativo
Fondo B	Flujo Negativo	Flujo Negativo
Fondo C	Flujo Positivo	Flujo Negativo
Fondo D	Flujo Positivo	Flujo Negativo
Fondo E	Flujo Positivo	Flujo Negativo
Fondo F	Flujo Negativo	Flujo Negativo
Fondo G	Flujo Positivo	Flujo Negativo
Fondo H	Flujo Positivo	Flujo Negativo
Fondo I	Flujo Positivo	Flujo Positivo
Fondo J	Flujo Positivo	Flujo Positivo

Tabla 12. Predicciones realizadas para el mes de noviembre con la herramienta implementada, fondos anonimizados

Resultados para los cuales se obtienen los siguientes valores de recall, que son las métricas que queremos maximizar, recordemos que la fórmula del recall es la siguiente:

$$Recall = \frac{TP}{TP + FN}$$

Etiqueta	Recall observado
Flujo Positivo	100%
Flujo Negativo	37,5%

Tabla 13. Recalls observados para FF.MM. Principal, noviembre 2023

Actuando bajo las predicciones e insights generados por el nuevo análisis, el ejecutivo de ventas fue capaz de generar una venta del Fondo D a un cliente institucional que no habría sido considerada bajo la herramienta antigua, resultando en un flujo adicional correspondiente a un 0,12% por sobre las ventas del mes, atribuible en parte al apoyo generado y la estrategia de venta basada en los insights de este.

10.3. Factores Cualitativos

Aparte de la reducción en tiempos de uso y de las predicciones realizadas, la presencia de los coeficientes del modelo para cada super sector presente en los datos permite generar una mayor cantidad de insights, ayudando a identificar las variables que impulsan las compras y ventas de los clientes a partir de sus pesos en el modelo y los resultados logrados, esto complementa los conocimientos del ejecutivo ya que permite identificar los patrones de valorizaciones de los clientes, que apreciaron durante el periodo estudiado y que castigaron, ayudándole a crear una mejor estrategia comercial personalizada y con sus tendencias en menta.

11. Discusión y Conclusiones

En conclusión, se cumple el primer objetivo específico relacionado al tiempo de procesado y análisis, destacando que la plataforma realiza sus cálculos con recursos cloud, por lo que el ejecutivo puede seguir llevando a cabo otras labores mientras se obtiene el análisis mensual o permitiéndole utilizar Excel para otros motivos al mismo tiempo. Lo anterior sumado a la reducción del uso de tiempo en un 66,6% es que se considera un éxito.

Por otra parte, el objetivo relacionado a la calidad de los insights no se alcanza a cumplir por una diferencia de 0,02% en respecto a la métrica propuesta a inicios del proyecto. Esto debido a que la venta realizada por el ejecutivo no cumplió con tener un tamaño promedio, resultando en este diferencial negativo. Sin embargo, se debe destacar no solo el funcionamiento del modelo en respecto a las predicciones realizadas, sino que también la situación económica y financiera en la que se encuentra el mundo. Este ha sido un año especialmente complejo, dos guerras se mantienen en curso, la inflación y el desempleo se mantienen altos, los ahorros reales se encuentran en puntos más bajos históricos.

No es de extrañar que los resultados enmarcados en este año no sean ejemplares, por lo que lo logrado se puede considerar suficiente para validar el proyecto como un éxito, esperando que se puedan generar mejores flujos a partir de *insights* del nuevo análisis cuando la economía y el país se encuentren en situaciones más estables para la inversión.

12. Referencias

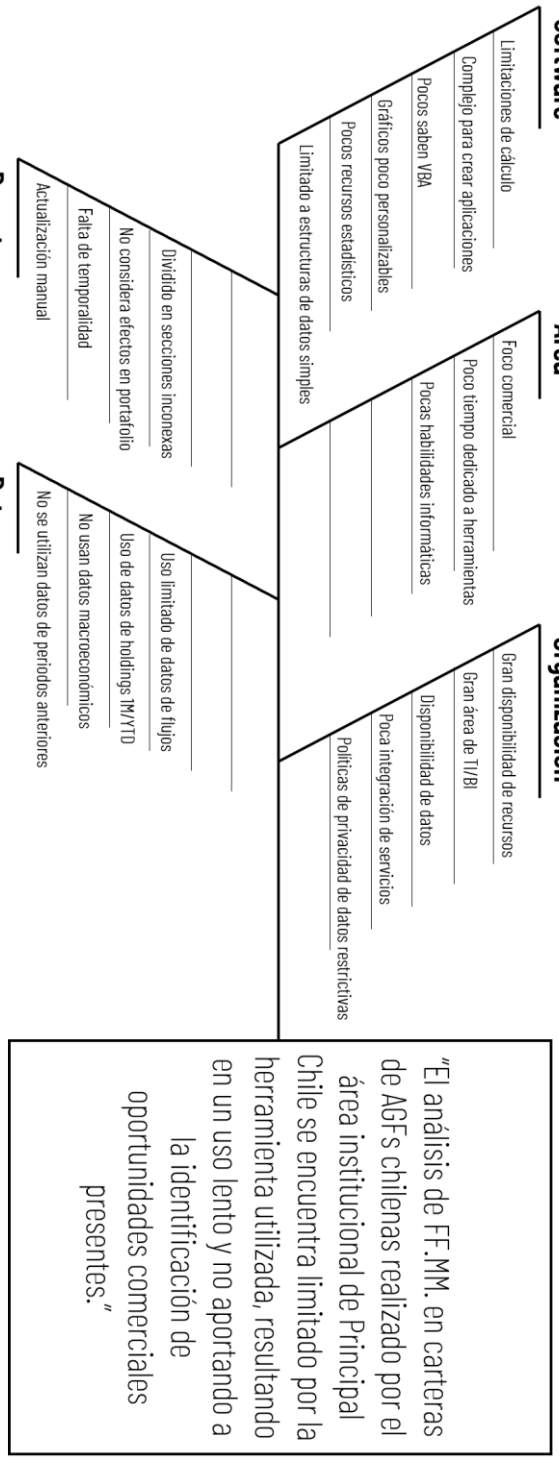
- R. Salman and V. Kecman, "Regression as classification," 2012 Proceedings of IEEE Southeastcon, Orlando, FL, USA, 2012, pp. 1-6, doi: 10.1109/SECon.2012.6196887.

- Andreas, Christopher & Faricha, Anifatul & Ulyah, Siti & Susanti, Rika & Mardhiana, Hawwin & Achirul Nanda, Muhammad & R., Firman. (2022). Comparison study using ARIMAX and VARX in cash flow forecasting. AIP Conference Proceedings. 2641. 030023. 10.1063/5.0118519.

- Aneesh Tony, Pradeep Kumar, Rohith Jefferson, Subramanian, "A Study of Demand and Sales Forecasting Model Using Machine Learning", Psychology and Education, 2021.

13. Anexos

13.1. Diagrama de Ishikawa



13.2. Regresores utilizados

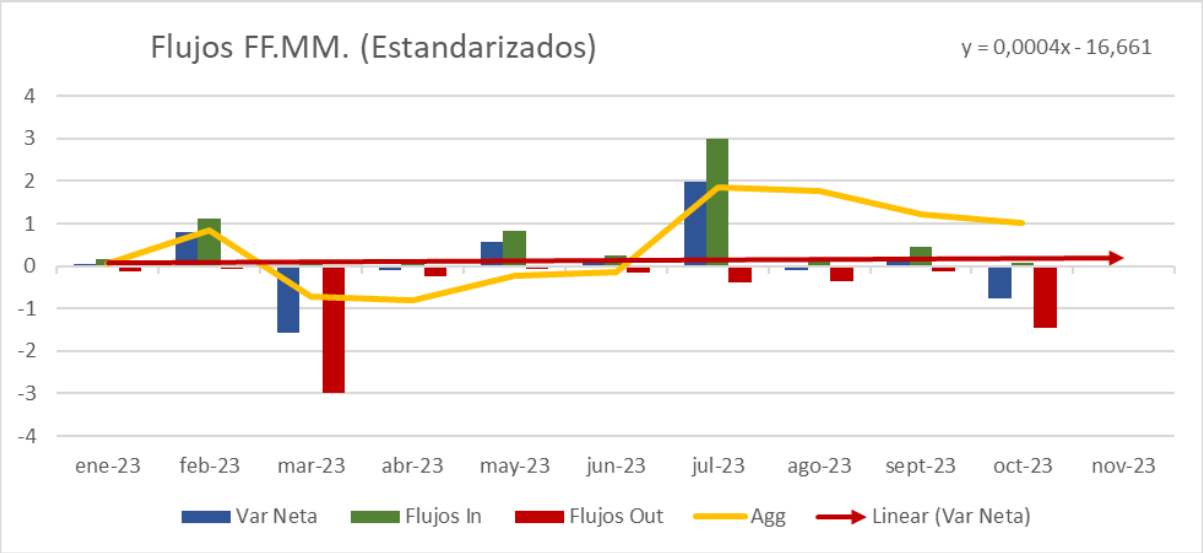
A continuación se muestra una enumeración con los regresores utilizados para entrenar al modelo predictivo, donde Ret son retornos, Vol son volatilidades y Var son variaciones porcentuales.

- Ret. Mes [1-6]
- Ret. Mes [7-12]
- Ret. Mes [13-18]
- Ret. Mes [19-24]
- Ret. Mes [25-30]
- Ret. Mes [31-36]
- Vol. Mes [1-6]
- Vol. Mes [7-12]
- Vol. Mes [13-18]
- Vol. Mes [19-24]
- Vol. Mes [25-30]
- Vol. Mes [31-36]
- Var. 1M dólar
- Var. 1Y dólar
- Var. 1M Imacec
- Var. 1Y Imacec
- Var. 1M Inflación US
- Var. 1Y Inflación US
- Var. 1M TPM
- Var. 1Y TPM
- Var. 1M TPM Fed
- Var. 1Y TPM Fed
- Var. 1M Brent
- Var. 1Y Brent
- Var. 1M Oro
- Var. 1Y Oro
- Var. 1M VIX
- Var. 1Y VIX

13.3. Tiempos de Uso de la Herramienta Antigua

	Veces por mes	**Media**	**Desviación estándar**	**Tiempo total promedio mensual**
Extracción y preparación de datos.	1	5 minutos	10 segundos	5 minutos
Análisis de datos	20	45 minutos	5 minutos	15 horas
Cantidad de refresco de datos por sesión	8	1 minuto	5 segundos	2 horas, 40 segundos

13.4. Flujos Netos, Positivos y Negativos Observados en el Año



13.5. Tabla Resumen de Performance Modelos Binomiales

Row Labels	-1 : Precision	-1 : Recall	-1 : F1 Score	1 : Precision	1 : Recall	1 : F1 Score	Accuracy
⊞Extra Tree Classifier	71,9%	62,0%	66,4%	37,7%	48,9%	42,1%	57,8%
⊞Extra Trees Classifier	71,4%	71,3%	71,2%	39,4%	39,4%	39,0%	61,3%
⊞KNN Classifier (k = 10)	69,5%	91,3%	78,8%	42,4%	14,2%	20,1%	67,0%
⊞Linear SVC	71,7%	55,7%	62,4%	36,4%	53,1%	42,6%	54,9%
⊞Logistic Regression	71,4%	54,9%	61,6%	36,1%	53,1%	42,4%	54,4%
⊞Passive Aggressive Classifier	64,2%	53,5%	52,7%	32,5%	49,8%	33,4%	52,4%
⊞Perceptron	63,5%	50,1%	50,2%	31,9%	52,7%	34,7%	50,9%
⊞Random Forest	71,4%	72,0%	71,5%	39,5%	38,6%	38,7%	61,6%
⊞SGD Classifier	66,0%	52,2%	52,6%	34,3%	51,8%	35,3%	52,1%
⊞SVC	72,7%	49,1%	56,1%	36,3%	59,6%	43,6%	52,3%
Grand Total	69,4%	61,2%	62,3%	36,6%	46,1%	37,2%	56,5%

13.6. Tabla Resumen de Performance Modelos Multinomiales

Row Labels	-1 : Precision	-1 : Recall	-1 : F1 Score	0 : Precision	0 : Recall	0 : F1 Score	1 : Precision	1 : Recall	1 : F1 Score	Accuracy
⊞Extra Tree Classifier	45,8%	43,1%	44,1%	54,5%	43,5%	48,1%	22,7%	37,1%	27,8%	42,3%
⊞Extra Trees Classifier	45,8%	43,3%	44,3%	52,6%	49,4%	50,7%	23,4%	29,4%	25,7%	43,7%
⊞KNN Classifier (k = 10)	42,0%	49,6%	45,0%	47,6%	52,5%	49,5%	24,2%	8,4%	11,7%	44,0%
⊞Linear SVC	41,4%	42,4%	41,2%	45,2%	48,8%	46,4%	20,9%	15,4%	16,8%	41,0%
⊞Logistic Regression	41,1%	38,3%	39,1%	45,1%	28,9%	34,7%	20,0%	40,2%	26,4%	34,6%
⊞Passive Aggressive Classifier	35,4%	36,4%	29,5%	40,1%	37,8%	32,0%	14,9%	28,7%	15,3%	35,8%
⊞Perceptron	34,2%	35,6%	28,6%	37,6%	36,6%	31,2%	15,2%	30,4%	15,9%	35,2%
⊞Random Forest	45,9%	43,6%	44,5%	52,5%	49,6%	50,8%	23,5%	29,1%	25,6%	43,9%
⊞SGD Classifier	39,6%	35,6%	31,3%	43,8%	42,4%	37,4%	17,1%	26,4%	16,4%	37,2%
⊞SVC	41,6%	35,3%	35,4%	43,6%	26,9%	29,9%	20,1%	45,7%	26,6%	33,6%
Grand Total	41,3%	40,3%	38,3%	46,3%	41,6%	41,1%	20,2%	29,1%	20,8%	39,1%

13.7. Matriz de Correlaciones de Variables Exógenas