



MAESTRÍA EN CIENCIA DE LOS DATOS Y ANALÍTICA

PROYECTO FINAL - MINERÍA DE DATOS PARA GRANDES  
VOLÚMENES DE INFORMACIÓN

*Desarrollo de algoritmos de machine learning para la rentabilización  
de portafolios de renta fija soberana basados en información  
transaccional del mercado*

**Estudiantes:**

Carlos Castro Marin

David Mejia Estrada

Oscar Javier Contreras Buitrago

**Profesor:**

Carlos Alzate

2023

# 1 Entendimiento del negocio

## 1.1 Determinación de objetivos del negocio

La predicción del comportamiento de mercados financieros siempre ha sido el objetivo principal de la industria de administración de activos, valiéndose históricamente de numerosas técnicas computacionales, estadísticas, de análisis técnico y fundamental para lograr múltiples objetivos como maximizar la rentabilidad de portafolios, reducir el riesgo, entre otros. Si bien este tipo de herramientas y estrategias pueden implementarse en distintos ámbitos y geografías, en el ambiente nacional y con datos disponibles de manera libre y actualizada nace la pregunta: *¿Es posible generar rentabilidad en un portafolio de renta fija soberana colombiana a través de modelos de machine learning que utilicen información transaccional del mercado de valores como input?*

Para dar respuesta a esta pregunta se propone como objetivo de este proyecto: desarrollar modelos de Machine Learning sobre información transaccional del mercado de valores de deuda pública colombiana, cuyo objetivo sea la generación de rentabilidad a través de compraventas en dicho mercado. Para lograr este cometido se plantean los siguientes objetivos específicos:

- Transformar utilizando Spark un dataset que contenga la información transaccional del mercado de renta fija soberana colombiana (operado en el Sistema Electrónico de Negociación SEN), para encontrar transformaciones de los datos que puedan ser de utilidad para algoritmos de machine learning.
- Utilizar modelos de machine learning sobre los datos con el objetivo de lograr mejores rentabilidades a través de pronósticos de movimiento de precios y curvas de tasas de interés. Utilizando para esto las librerías de machine learning propias de Spark.
- Disponibilizar tanto la transformación de los datos como el entrenamiento y testeo de los modelos de machine learning en un repositorio en Github, que permita una consulta posterior, ajuste de los algoritmos implementados o puesta en vivo como herramienta financiera de los mismos

## 1.2 Evaluación de la situación Estado del arte

El aprendizaje automático ha sido una solución común para administración de portafolios en los mercados de capitales, donde es ampliamente conocido que los instrumentos financieros en los mercados financieros siguen un movimiento browniano geométrico, que es un proceso aleatorio markoviano Merton (1973) Black and Scholes (1973). De esta manera, las aplicaciones de aprendizaje automático en los mercados financieros continúan siendo una rama de investigación muy importante de la inteligencia artificial y que se encuentra en continuo desarrollo. Los comienzos del aprendizaje automático en los mercados financieros están relacionados con el surgimiento de las redes neuronales a fines del siglo pasado, donde se trataron aplicaciones sobre la fijación de precios de opciones y la estructura temporal de las tasas de interés, entre otros, con resultados iniciales satisfactorios Refenes et al. (1997) y Zimmermann et al. (2000).

Los algoritmos de aprendizaje automático comenzaron a utilizarse para aplicaciones de negociación en los mercados financieros aproximadamente por la misma época, cuando Moody y Saffell Moody and Saffell (1998) demostraron que se puede entrenar un algoritmo de aprendizaje automático recurrente para operar con carteras de acciones optimizando el ratio de Sharpe, que es una medida de la relación entre la rentabilidad total obtenida y la

cantidad de riesgo de mercado asumido.

En la última década, con la aparición y el auge del aprendizaje de máquina profundo, también han surgido nuevos enfoques de aprendizaje automático profundo, con nuevas aplicaciones en diferentes mercados. Los autores han escrito especialmente sobre aplicaciones en el mercado de valores con enfoques diferentes, donde las estrategias de negociación de acciones fueron el centro de esos estudios Wu et al. (2020), Li et al. (2020), Park et al. (2020), Carta et al. (2021), y en los cuales el rendimiento mejoró en relación con los métodos de redes neuronales básicas, así como los métodos convencionales de “buy and hold”, mostrando buenos resultados incluso durante la crisis del covid-19 Yang et al. (2020). Además, aparecieron aplicaciones de aprendizaje automático para el mercado de divisas Carapuço et al. (2018) e incluso para la negociación en los mercados de energía Chen and Su (2018), dos mercados financieros con grandes diferencias respecto al ampliamente estudiado mercado accionario.

Por otro lado, las investigaciones de aprendizaje automático nunca tuvieron especial interés en la negociación o gestión y optimización de carteras de instrumentos de renta fija, por lo que existe un gran vacío en la literatura sobre esos temas Nunes (2022). En los pocos estudios existentes al respecto, el enfoque se centraba en el modelado de la curva de rendimiento, la predicción de su figura y patrones, movimientos y, en algunos casos, la predicción de crisis financieras inminentes, todo utilizando Máquinas de Soportes Vectoriales y Análisis de Componentes Principales, entre otros métodos Gogas et al. (2015) y Dixon et al. (2020). En un intento por llenar esta brecha, Nunes (2022), en su tesis doctoral, encontró que los algoritmos de aprendizaje por refuerzo DDPG tenían un buen rendimiento en la negociación de ETF (fondos cotizados) de renta fija. Sin embargo, los ETF pueden entenderse como instrumentos de renta variable pese a que su subyacente pueda ser perfectamente uno o varios instrumentos de renta fija. Así, se concluye que el uso del aprendizaje de máquina para la negociación de valores de renta fija no es común en la literatura.

### **1.3 Determinación de metas de la minería de datos**

El principal objetivo de esta etapa, par el proyecto en cuestión, es encontrar las variables y transformaciones necesarias para obtener un Dataset basado en en información transaccional del mercado de valores de deuda pública colombiana. Posteriormente este Dataset, producto de las operaciones definidas, sea apto para alimentar distintos modelos de Machine Learning cuyo objetivo sea la generación de rentabilidad a través de compraventas. Esto apunta directamente a l logro el objetivo principal del proyecto planteado.

### **1.4 Generación de un plan de proyecto**

Las fechas propuestas y etapas a desarrollar en el proyecto se encuentran resumidas en el diagrama de Gantt de la figura 3.

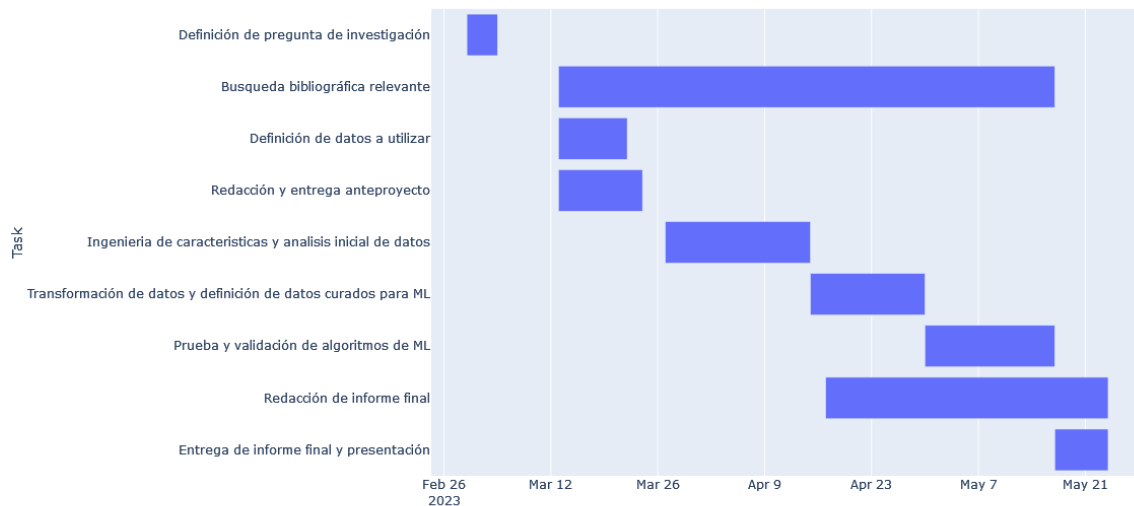


Figura 1: Plan de desarrollo del proyecto - Diagrama de Gantt

El desarrollo del plan y sus resultados son mostrados a continuación.

## 2 Entendimiento de los datos

### 2.1 Recolección de los datos

Los datos recopilados corresponden a los cierres individuales de operaciones de compraventa o simultáneas registrados en el mercado de renta fija soberana de títulos de tesorería (TES) a través del SEN, administrado por el Banco de la República. Los datos son públicos y se pueden encontrar en la página web del Banco de la República encontrados en BanRep (2023). Cada uno de los registros, corresponde, entonces, a una transacción pactada entre dos contrapartes, y puede estar registrada en dos ruedas:

- La rueda de simultáneas, las cuales son operaciones de liquidez entre dos contrapartes, en donde el acreedor recibe un título, en este caso, TES, como garantía del préstamo, y cuyo plazo debe ser superior a 3 días hábiles.
- La rueda de contado, en donde se encuentran las compraventas efectivas entre las dos contrapartes sobre TES a un precio determinado por ellas, con cumplimiento el mismo día

### 2.2 Descripción de los datos

La información de estas transacciones puede ser adquirida de manera mensual, por esta razón, los datasets mensuales deben ser consolidados en un solo Dataset maestro. La información en cada columna, así como su significado práctico se encuentran a continuación.

- **Fecha de cierre (aammdd):** Fecha en la cual la operación fue pactada
- **Hora de cierre (hhmmss):** Hora en la cual la operación fue pactada
- **Rueda:** Determina la rueda en la cual se dio la negociación, y por tanto, el tipo de operación

- **Plazo:** Aplica para rueda de simultáneas, es el tiempo en días hábiles en el cual el acreedor devolverá la garantía y el deudor devolverá el dinero prestado más los intereses causados
- **Instrumento:** Se refiere al identificador o ticker del instrumento TES objeto de la operación.
- **Tasa/Precio:** Para la rueda de simultáneas, es la tasa de interés pactada para la operación de liquidez. Para la rueda de contado, es el precio limpio del TES pactado en la compraventa
- **Tasa equivalente:** Para la rueda de contado, representa la Yield To Maturity (YTM) pactada en la operación de compraventa. Para la rueda de simultáneas es la misma tasa de la columna anterior
- **Valor nominal:** Es el valor nominal del TES objeto de la operación. Dependiendo del título TES puede estar dado en pesos colombianos o en Unidades de Valor Real (UVR)
- **Contravalor:** Es el valor de giro en pesos de la operación
- **Valor nominal del colateral:** Solo para simultáneas. Valor nominal en pesos o UVR del colateral de la operación. Si es rueda de contado va en cero
- **Precio colateral:** Es el precio del TES utilizado como colateral en la operación
- **Tasa equivalente colateral:** Es la YTM equivalente al precio del colateral en la operación

Ya que uno de los datos a utilizar para el desarrollo del proyecto son los **Instrumentos** ahondaremos un poco en su nomenclatura. Un Instrumento tiene la siguiente estructura: **TFIT16240724**; donde el nemotécnico quiere decir, según sus caracteres:

- 1: TES Clase B.
- 2-4: Tipo de Tasa y Amortización
- 5-6: Años comprendidos entre la emisión y el vencimiento
- 7-12: Fecha de vencimiento en DDMMAA

### 2.3 Exploración de los datos y verificación de la calidad de estos

La exploración de la Data obtenida, su calidad y las transformaciones aplicadas a estos para su ingesta en los modelos de Machine Learning serán expuestos en etapas posteriores. Esto para dar foco al análisis de los datos tratados y modificados, mas allá que a las fuentes originales consultadas

## 3 Metodología de investigación

Para el desarrollo del proyecto en cuestión se propone utilizar la metodología CRISP-DM (*Cross-Industry Standard Process for Data Mining*) de IBM IBM (2018), este es una metodología probada industrialmente para guiar los procesos de minería de datos cuya estructura se puede apreciar en la figura 2

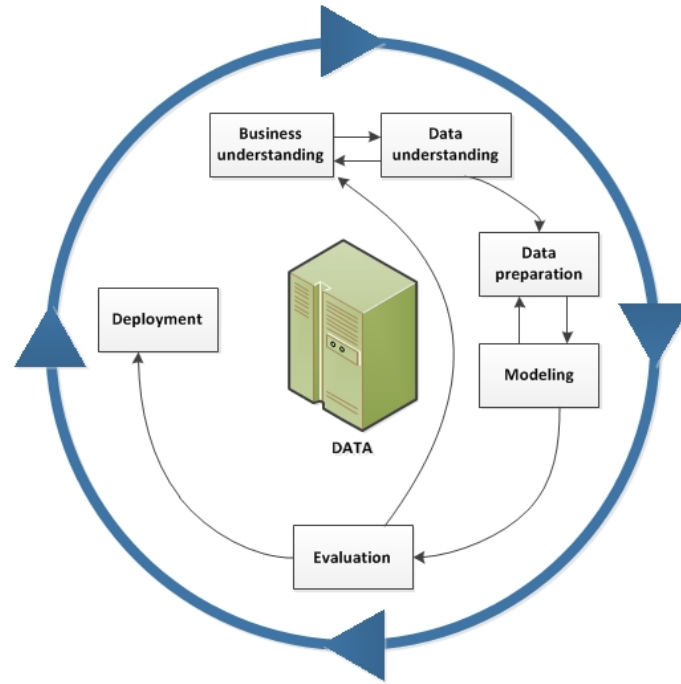


Figura 2: Ciclo de vida de la minería de datos IBM (2018)

## 4 Preparación de los datos

Para la preparación de los datos, en el script *consolidar\_data.py*, se implementan diversas técnicas de preparación de datos con el fin de procesar archivos de Excel que contienen información sobre operaciones financieras, tomados de BanRep (2023). A continuación, se describe de manera amplia el enfoque de preparación de datos utilizado en el código:

1. **Exploración y comprensión de los datos:** Se realiza una exploración inicial para identificar los tipos de datos, la estructura de los archivos de Excel y las características relevantes que se encuentran en los mismos para su posterior tratamiento.
2. **Limpieza de datos:** Se llevan a cabo acciones para eliminar o corregir datos incorrectos, incompletos o duplicados, lo que ayuda a mejorar la calidad de los datos y a garantizar la integridad de los análisis posteriores. Los principales errores hallados corresponden a las fechas y horas de las transacciones.
3. **Filtrado y selección de datos:** Se aplican filtros y condiciones específicas para identificar y mantener únicamente los datos que son de interés para el estudio en cuestión. En específico, no deseábamos títulos diferentes a TES en tasa fija, o que se negociaran en rueda de simultáneas, pues solo la rueda de contado es de nuestro interés, esto visto desde el conocimiento experto del negocio.
4. **Integración y consolidación de datos:** La integración de datos implica combinar y consolidar los datos tratados previamente en un formato coherente y estructurado, lo que facilita su análisis y explotación. Se consolidan todos los datos en un solo Dataframe, indexado por la concatenación de fechas y horas de transacción y con la información de los Instrumentos considerados.
5. **Guardado y documentación de los datos:** Una vez que los datos han sido procesados y preparados, es importante guardarlos en un formato adecuado para su posterior uso. El script guarda los datos en un archivo .csv, pero también existe

la posibilidad de guardarlos en un `.parquet`. Esto dependerá del espacio disponible para almacenar la Data resultante y el uso que se plantee para esta.

En el código presentado se aplicaron las técnicas de limpieza, transformación, filtrado y consolidación para garantizar que los datos estén en un estado adecuado para el análisis de estos desde una perspectiva bursátil. Estas técnicas permitieron mejorar la calidad y la consistencia de los datos, lo que a su vez contribuyó a la obtención de resultados precisos y confiables en el análisis posterior y con relevancia financiera. Esta última afirmación se realiza pensando en la necesidad de que este tipo de pre-procesamientos debe basarse en una mezcla de técnicas de análisis de datos y el conocimiento experto del negocio, pues cualquier falla en alguna de estas puede impactar el resultado final del mismo.

A partir de este punto, en el *Notebook Proyecto\_Mineria\_de\_Datos\_Castro\_Contreras\_Mejia.ipynb* de jupyter que se encuentra en el repositorio, realiza la preaparación los datos a través de transformaciones en spark con el objetivo de generar un nuevo Dataframe, con características bastante diferentes, y que servirá de input para los modelos de machine learning. Seguimos los siguientes pasos:

1. Eliminamos las siguientes columnas que no utilizaremos para el procesamiento:
  - 'SESION/RUEDA', pues solo trae como dato el string 'CONH', nombre de la rueda de contado.
  - 'TASA/ PRECIO', puesto que no buscamos hacer predicciones sobre el precio, sino por la Tasa al Vencimiento, que está en otra columna. Esas dos columnas están relacionadas inversamente, entonces no aporta valor a la predicción el mantener los precios en el dataset.
  - 'VR. NOMINAL', pues no interesa el tamaño de la transacción, únicamente la tasa al vencimiento.
  - 'CONTRAVALOR', pues es una columna que solo trae datos cuando se trata de la rueda de negociación de simultáneas.
2. Se concatenan las columnas de FECHA y HORA en una sola columna en formato datetime para posteriores etapas de procesamiento.
3. Filtrado del Dataset para incluir únicamente INSTRUMENTOS del tipo TFIT
4. Se modifica la temporalidad de los datos enviándolos a la hora inferior en cada registro. El objetivo es trabajar las predicciones en intervalos de una hora, logrando agrupar los datos y representarlos con el dato de cierre de dicha hora, siguiendo las prácticas comunes del sector financiero.
5. Se genera una ventana horaria para la data, además se genera la nueva data *last\_yield* correspondiente al último valor de **YTM** en cada espacio horario por INSTRUMENTO en el Dataset.
6. Se filtran títulos relevantes para el análisis, en este caso los títulos definidos corresponden a títulos de alta liquidez y con tiempos de transacción similares. Los demás títulos no son reconocidos como de alta liquidez, lo que implica que pueden llegar a generar bastantes problemas en los modelos de machine learning por la alta prevalencia de valores Null y por su poca profundidad en el mercado de valores.
7. Por último se realiza el pivoteo de la data por los INSTRUMENTOS seleccionados, esto genera una nueva estructura de datos que permite ver los valores de YTM para cada Instrumento seleccionada en cada intervalo de tiempo del Dataset original.

Con estas transformaciones obtenemos la tabla final que será utilizada para el modelado con machine learning implementado en spark con la librería MLib Meng et al. (2015) en el mismo notebook *Proyecto\_Mineria\_de\_Datos\_Castro\_Contreras\_Mejia.ipynb* utilizado en las transformaciones mencionadas.

## 5 Modelado

Para el proceso de modelamiento se usaron diferentes herramientas de la librería de machine learning para Apache Spark MLib Meng et al. (2015). La modulación se encuentra en el mismo notebook *Proyecto\_Mineria\_de\_Datos\_Castro\_Contreras\_Mejia.ipynb*. En este apartado, el DataFrame de Spark es utilizado para correr dos modelos de regresión de Machine Learning diferentes: Random Forest y Gradient Boosting Tree.

Inicialmente se instalan librerías de Pyspark necesarias para la estructuración de los datos a ser entregados a los modelos, así como los modelos en como tal. Para ambos modelos, los datos son separados 70% - 30%, en training y testing, respectivamente.

El primer modelo planteado es el Random Forest. Con un RMSE de 0.33 en las predicciones puntuales, y una composición de la predicción que utiliza la curva de rendimientos de la siguiente manera: 41% en TES del 26, 37% en TES del 24 y 22% en TES del 28. El RMSE muestra que la desviación media sobre la predicción puntual se ubica sobre los 10 puntos básicos, cifra que, si bien no es muy alta teniendo en cuenta las volatilidades que puede experimentar la tasa de estos títulos durante un día, si puede implicar pérdidas significativas en el largo plazo.

En ese sentido, procedemos a evaluar el Accuracy sobre la predicción de la dirección de la tasa. Al evaluar en términos de Accuracy sobre la capacidad del modelo, no de predecir el valor puntual de la tasa de interés de mercado, sino de su dirección para el día siguiente, encontramos que esta medida de desempeño tiene un valor aproximado de 0.6, que, en el contexto financiero de bursátil, es un gran resultado, y que abre la posibilidad de que el modelo esté cumpliendo con su objetivo de generar rentabilidad.

En ese sentido, evaluamos para cada día la cantidad de puntos básicos que la estrategia es general al tomar posición. El modelo toma posición de compra cuando la predicción de la dirección sea a la baja, y de venta, cuando la predicción sea que la tasa de interés del título se incrementará. Por último, teniendo en cuenta la posición tomada, calculamos la cantidad de puntos básicos que la estrategia obtuvo, sean de utilidad o pérdida. Es importante recordar que, a mayor tasa de interés de mercado, menor será el precio de mercado del título. Por lo tanto, en trading de renta fija, se genera incremento en el precio cuando la tasa de interés cae  $X$  puntos básicos, y se genera disminución del precio cuando la tasa de interés sube.

Repetimos el proceso para el modelo de Gradient Boosting, en donde el Gradient Boosted Tree también toma en cuenta una muestra para entrenamiento del 70% del DataFrame total. En este caso, los puntos básicos medios en los que la predicción se desvía del valor real es de 6 puntos básicos por ventana de tiempo. La evaluación del Accuracy sobre la dirección también mejora, llegando a aproximadamente 65%. Los resultados comparando ambos algoritmos son explicados en un apartado posterior.



## 6 Evaluación

En la siguiente tabla se muestra un resumen de los resultados de cada uno de los modelos, en donde reiteramos que si bien el RMSE puede considerarse bajo para lo que es la volatilidad de los títulos en el mercado, no es suficientemente bueno como para intentar realizar trading buscando predicciones puntuales en intraday. En cambio, el Accuracy es bastante bueno, entendiendo que para este tipo de problemas, se busca que los accuracies superen o estén en cualquier punto entre 0% y 45% o 55% y 100%.

|                                 | Random Forest Regressor | GBTRegressor |
|---------------------------------|-------------------------|--------------|
| <b>RMSE</b>                     | 0.3304                  | 0.2491       |
| <b>Accuracy *</b>               | 0.5899                  | 0.6572       |
| <b>Puntos basicos ganados</b>   | 4299.9                  | 5883.5       |
| <b>Importancia TFIT15260826</b> | 0.4133                  | 0.3848       |
| <b>Importancia TFIT16240724</b> | 0.3684                  | 0.6151       |
| <b>Importancia TFIT16280428</b> | 0.2182                  | 0.0000       |

Al tomar decisiones de inversión tomando en cuenta el accuracy de los modelos, obtenemos entonces una utilidad total de 4.299,99 puntos básicos y 5.883,50 puntos básicos, para Random Forest y para el Gradient Boosted Tree, respectivamente. Es interesante anotar que el modelo de Gradient Boosting logra mejores utilidades sin tomar en cuenta la información de los TES del 28.

Después de concluir que ambos modelos pueden generar una gran cantidad de puntos básicos, es decir, de utilidad, si se quiere, se puede estimar dicha utilidad en términos monetarios multiplicando esa cantidad de puntos básicos por un DV01 promedio para los TES con vencimiento en julio de 2024. Este DV01 promedio es de 350 mil pesos para un portafolio de mil millones de pesos (Se puede escalar dependiendo del tamaño del portafolio).

Recordemos que el DV01, o *dollar duration*, es una medida de sensibilidad de portafolios de renta fija en términos monetarios. Para las prácticas del mercado financiero colombiano, el DV01 es la utilidad en pesos que se genera en un portafolio de renta fija de x activos por la variación en un punto básico en la tasa de interés.

## 7 Despliegue

### 7.1 Montaje de información en GitHub

El código generado en las etapas de procesamiento y modelamiento se encuentran disponibles en el repositorio de GitHub siguiente:

[https://github.com/dmej98/Repo\\_GrandesVoldeInfo](https://github.com/dmej98/Repo_GrandesVoldeInfo).

### 7.2 Seguimiento del proyecto y lecciones aprendidas

En el desarrollo real del proyecto se observa un offset en algunos de los puntos desarrollados, los principales cambios se encuentran en que: Las etapas iniciales, principalmente la búsqueda bibliográfica, fueron realizadas en la parte inicial y no se extendieron a la parte final del proyecto como se planeaba. El otro gran cambio se dio en que las etapas de Ingeniería de características, transformación de los datos y modelado no se dieron de manera secuencial; se dieron casi de manera simultanea pues en el proceso de realizar cada

etapa se realizaron iteraciones que necesitaron volver a etapas previas.

Como lecciones aprendidas, como en esta clase de proyectos, consideramos necesario considerar la ciclicidad propia de la metodología CRISP-DM y los proyectos de datos, donde los procesos no son lineales y por esto el diagrama de proyecto debe reflejarlo, adicionalmente, consideramos necearía realizar sesiones de avance internas pues la parte final del proyecto fue recargada en las labores de programación y no de generación del informe y presentación como se había planteado.

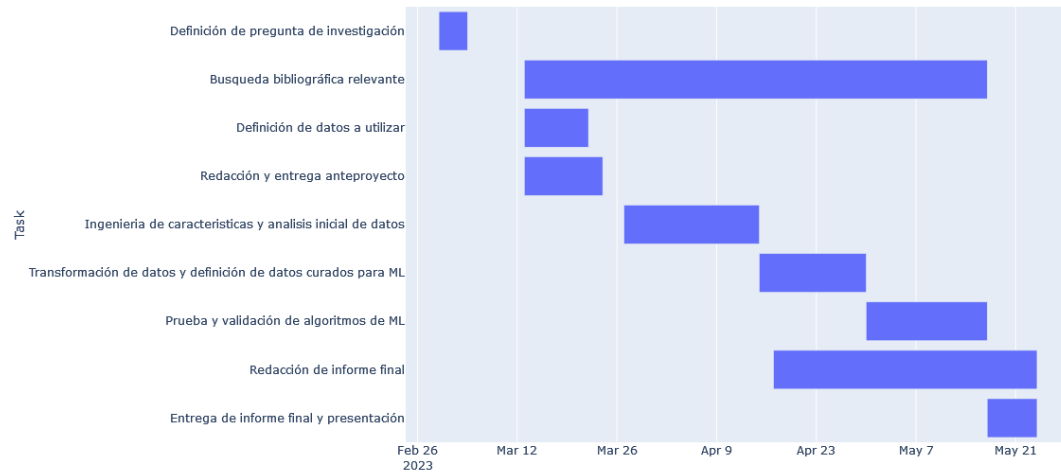


Figura 3: Plan de desarrollo del proyecto - Diagrama de Gantt

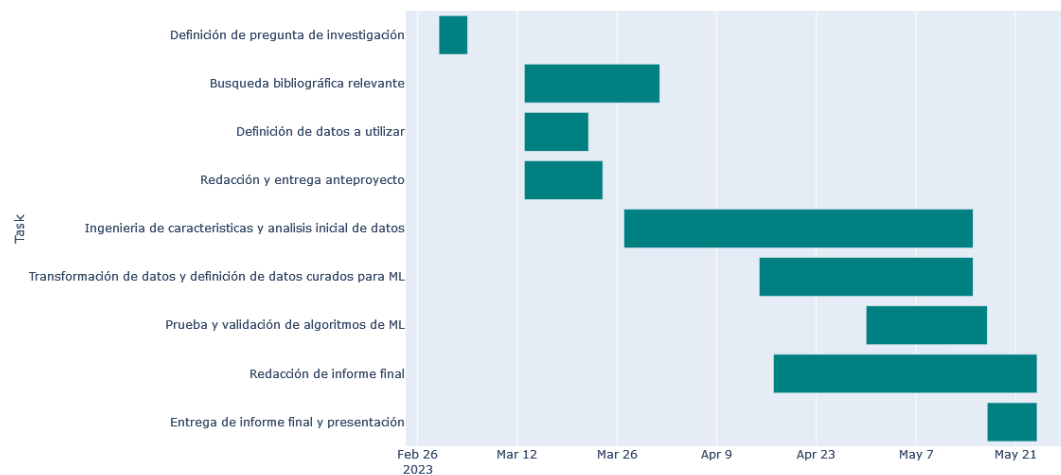


Figura 4: Ejecución real del proyecto - Diagrama de Gantt

### 7.3 Conclusiones y trabajos futuros

Sobre nuestra pregunta de investigación, concluimos que es posible generar rentabilidad en este tipo de portafolios utilizando modelos de machine learning usando información transaccional, pese a que la cantidad de información en grandes cantidades pueda parecer una limitante para el procesamiento de los datos. Esto se logra con la utilización de her-

ramientas que permiten realizar transformaciones y acciones sobre esos grandes volúmenes de información de manera ágil.

En ese sentido, concluimos sobre nuestros objetivos específicos, lo siguiente:

- Apache Spark es una herramienta que permite trabajar con agilidad sobre grandes volúmenes de información transaccional del SEN. Desde un gran Data Set, pudimos incorporar varias transformaciones, llegando a un Data Set del tamaño justo que nuestros modelos de machine learning necesitaban.
- Con Spark, dado el gran volumen de datos, es también posible implementar modelos de machine learning con las propias librerías que este incorpora.

Los resultados positivos sobre esta implementación son bastante motivadores, y nos impulsan a pensar sobre trabajos futuros que pueden, inclusive, ser más exigentes en cuanto a volumen de información. Por enumerar algunos:

- Incorporar un análisis de Bid-Ask spread sobre la base de datos. Además, con datos de profundidad. Para esto es necesario tener acceso a software especializado del mercado de valores, como MasterTrader y SEN. Entrenar a los modelos con Bid-Ask spread y no con la tasa de interés puede ayudar a encontrar más precisión y a mejorar la metodología de evaluación y backtesting. Sin embargo, la base de datos se podría multiplicar en tamaño por varias veces, pues existen siempre más puntas que cierres efectivos de operaciones en el mercado. Un reto para Spark.
- Incorporar los modelos en streaming, como un servicio en la nube que permita a los traders disponer de las predicciones en todo momento.

## 8 Implicaciones éticas

Los modelos de Machine Learning sobre el mercado de renta fija soberana colombiana incorporados en este trabajo buscan la generación sostenida de rentabilidad en el marco temporal evaluado utilizando como mecanismo para esto la toma de decisiones financieras de inversión o desinversión, que, para el caso del mercado de valores colombiano, deben ser tomadas bajo criterio profesional y bajo vigilancia de entes reguladores como la Superintendencia Financiera de Colombia (SFC), y entes autorreguladores como el Autorregulador del Mercado de Valores (AMV).

Es importante siempre resaltar que, si bien ciertos modelos de Machine Learning pueden llegar a presentarse prometedores según la rentabilidad que entregaron, esto no garantiza que la rentabilidad en el futuro sea sostenida. Las rentabilidades pueden ser variables dependiendo de los niveles de riesgo de mercado, liquidez y contraparte asumidos.

## 9 Aspectos legales y comerciales

Ningún elemento de este trabajo debe ser considerado como una recomendación de inversión ya que el estudio sólo incorpora aspectos estructurales del mercado de renta fija soberana colombiana que son relevantes para administradores de portafolios con criterios profesionales de inversión, por tanto, este trabajo es solo una aproximación académica a las inversiones en renta fija. Toda decisión de inversión o desinversión deberá estar dentro del marco ético y regulatorio de la AMV y SFC, así como de la legislación vigente.

En ese orden de ideas, las actividades de trading con recursos de terceros o la asesoría financiera, deberán estar previamente autorizadas por las anteriores entidades. Las actividades relacionadas al trading de posición propia no requieren autorización de las autoridades, sin embargo, el trading o la compraventa de activos financieros se considera como una actividad comercial que tendrá vigilancia bajo la normatividad colombiana.

En cuanto a los datos, no se reconocen asuntos éticos relevantes, esto debido a que la información sobre el mercado público de valores de renta fija está a disposición del público en general y los cierres de las transacciones son publicados por el Banco de la República de manera mensual en su página web.

## Referencias

- BanRep (2023). Cierres diarios. <https://www.banrep.gov.co/es/sen-puntuales-cierres>. Recuperado el 10 de mayo de 2023.
- Black, F. and Scholes, M. (1973). The pricing of options and corporate liabilities. *Journal of political economy*, 81(3):637–654.
- Carapuço, J., Neves, R., and Horta, N. (2018). Reinforcement learning applied to forex trading. *Applied Soft Computing*, 73:783–794.
- Carta, S., Ferreira, A., S. Podda, A., R. Recupero, D., and Sannai, A. (2021). Multi-dqn: An ensemble of deep q-learning agents for stock market forecasting. *Expert Systems with Applications*, 164:113820.
- Chen, T. and Su, W. (2018). Local energy trading behavior modeling with deep reinforcement learning. *IEEE Access*, 6:62806–62814.
- Dixon, M., Halperin, I., and Bilokon, P. (2020). *Machine Learning in Finance: From Theory to Practice*.
- Gogas, P., Papadimitriou, T., Matthaiou, M., and Chrysanthidou, E. (2015). Yield curve and recession forecasting in a machine learning framework. *Computational Economics*, 45:635–645.
- IBM (2018). Crisp-dm help overview. <https://www.ibm.com/docs/en/spss-modeler/18.2.0?topic=dm-crisp-help-overview>. Recuperado el 10 de mayo de 2023.
- Li, Y., Ni, P., and Chang, V. (2020). Application of deep reinforcement learning in stock trading strategies and stock forecasting. *Computing*, 102(6):1305–1322.
- Meng, X., Bradley, J. K., Yavuz, B., Sparks, E. R., Venkataraman, S., Liu, D., Freeman, J., Tsai, D. B., Amde, M., Owen, S., Xin, D., Xin, R., Franklin, M. J., Zadeh, R., Zaharia, M., and Talwalkar, A. (2015). Mllib: Machine learning in apache spark. *CoRR*, abs/1505.06807.
- Merton, R. C. (1973). Theory of rational option pricing. *The Bell Journal of Economics and Management Science*, 4(1):141–183.
- Moody, J. and Saffell, M. (1998). Reinforcement learning for trading. In Kearns, M., Solla, S., and Cohn, D., editors, *Advances in Neural Information Processing Systems*, volume 11. MIT Press.
- Nunes, M. (2022). *Machine learning in fixed income markets: forecasting and portfolio management*. PhD thesis, University of Southampton.
- Park, H., Sim, M. K., and Choi, D. G. (2020). An intelligent financial portfolio trading strategy using deep q-learning. *Expert Systems with Applications*, 158:113573.
- Refenes, A.-P. N., Burgess, A. N., and Bentz, Y. (1997). Neural networks in financial engineering: a study in methodology. *IEEE transactions on neural networks*, 8 6:1222–1267.
- Wu, X., Chen, H., Wang, J., Troiano, L., Loia, V., and Fujita, H. (2020). Adaptive stock trading strategies with deep reinforcement learning methods. *Information Sciences*, 538:142–158.

- Yang, H., Liu, X.-Y., Zhong, S., and Walid, A. (2020). Deep reinforcement learning for automated stock trading: an ensemble strategy. pages 1–8.
- Zimmermann, H., Neuneier, R., and Grothmann, R. (2000). Modeling of the german yield curve by error correction neural networks. pages 81–99.