Reporte

Especificación ML

Estimación de retiros netos para la administración de recursos de terceros

Juan Esteban Calderón Gómez

Universidad Nacional De Colombia

Contenido

El presente reporte tiene como finalidad presentar la especificación del modelo de aprendizaje de maquina para el problema de interés para trabajar

I. Motivación del proyecto de aprendizaje de máquina.

El problema especifico de aprendizaje de máquina, esta enfocado a la necesidad de predecir el comportamiento de una serie de tiempo, en este caso puntual se espera poder predecir la dinámica de retiros por parte de los clientes de una entidad financiera que administra fondos voluntarios de pensiones (FVP) de tal forma que, al conocer, previamente la cantidad de recursos que un grupo de clientes van a retirar de los fondos, se pueda realizar una mejor gestión de la liquidez (recursos disponibles de forma rápida e.g. cuentas de ahorros).

El hecho de predecir los mencionados retiros se articula con el problema formulado de forma directa, si se cuenta con un modelo de alta precisión, es de esperarse que se resuelva el planteamiento realizado y sobre todo que al tener un alto nivel de confiabilidad sea efectivamente usado en el día a día de la administración de portafolios.

II. Definición del problema.

De acuerdo con lo mencionado previamente, la salida del modelo es un número real $\mathcal{R}$ que representa el retiro neto probable a 1 y 10 días de cada uno de los FVP administrados por Fiduciaria Davivienda, el cual puede ser expresado indistintamente como un valor en dinero o como un porcentaje del valor del portafolio.

Los datos que permitirán predecir la salida mencionada anteriormente provienen de información disponible públicamente que actualiza la Superintendencia Financiera de Colombia, los cuales en términos generales tienen entre otras las siguientes características:

- Se les puede establecer como regla general un 70% de datos de entrenamiento y un 30% de testeo.
- Hay datos que son sesgados pero que tienen una explicación por ejemplo un retiro neto de -100% es común cuando comercialmente se lanzan alternativas al mercado que son a plazo, es decir después de x meses el portafolio se liquida y los recursos son devueltos a los clientes, pero no representa que necesariamente todos los clientes, en un mismo día han retirado la totalidad de sus recursos
- Por disposición de la Superintendencia estos datos se deben presentar de forma diaria por parte de las entidades, sin embargo, en la práctica se suelen ver algunos días de retraso. Si se trabajara exclusivamente con la información propia de la entidad, estos datos se actualizarían con certeza diariamente.

En línea con lo anterior, existe un factor que de desde una perspectiva de negocio podría ser aquel que pueda afectar de forma más especial el movimiento de la variable que se busca explicar, y es el retorno del portafolio, el cual podría esperarse que a medida que se presenten desvalorizaciones más negativas mayor es el nivel de retiros por parte de los clientes, de igual forma el tamaño del portafolio es una variable a tener en

cuenta, ya que es posible que para fondos más grandes se tenga frecuencia alta pero severidad baja, a diferencia de fondos pequeños donde se podría tener una alta concentración de recursos de un solo cliente que afectaría si retira todo su dinero.

De igual forma, dado que se ha evidenciado que las entidades que ofrecen este tipo de productos posiblemente conformen una industria homogénea, los demás fondos que tienen una dinámica similar, podrían ser una fuente de información que permita explicar la variable de interés.

III. Medidas de Desempeño.

En el caso del problema presentado, por tratarse de un grupo de series de tiempo, se trabajará inicialmente con el uso del RMSE el cual permite realizar una comparación entre los diferentes modelos que se puedan llegar a probar para los datos obtenidos.

Adicionalmente, y como una forma de reconocer que pueden existir estimaciones erróneas que tienen un impacto mayor que otras, es posible establecer un punto de validación de la calidad de la estimación de los modelos. En ese sentido en la tabla 1 se presentan las zonas de estimación de los modelos

		Observación	
		Retiro	Aporte
Estimación	Retiro	True Positive	False Positive
	Aporte	**False Negative**	True Negative

Tabla 1: zonas de estimación

En general el accuracy se calcula sumando el número de observaciones de la zona verde sobre el total de la suma de la tabla. Sin embargo en el caso de este modelo la zona roja es especialmente importante, ya que el propósito está centrado en estimar los retiros por parte de los clientes, y de cara a la dinámica del negocio tiene una implicación especial, el hecho de estimar un aporte cuando en realidad se va a presentar un retiro.

En línea con lo presentado previamente y por efecto de la ya mencionada lógica del negocio para el cual se construye el modelo, se determina un indicador de desempeño agrupado entre RMSE y los datos no estimados como falsos negativos.

Por ello el mejor modelo seleccionado será aquel que ocupe el mejor lugar ponderado entre el ranking por RMSE (20%) y el ranking por el porcentaje de estimaciones no catalogadas como falsos negativos (80%), dentro de la totalidad de los modelos estimados.

IV. Partición de los Datos.

Al tratarse de una serie de tiempo se trabaja con una partición estándar de 70% (datos de entrenamiento) y 30% (datos de testeo), de tal forma que la ventana de tiempo va siendo corrida para evaluar si la calibración diaria de los modelos es adecuada para estimar el dato siguiente en la serie.

V. Línea de Tiempo.

Los primeros resultados del presente estudio deben evidenciarse a más tardar durante la tercera semana del mes de Agosto de 2021, sin embargo el proceso de implementación al interior de la organización se espera tener hacia el mes de Diciembre de 2021 bajo una marcha blanca, para así tener una implementación limpia en el mes de marzo de 2022.

VI. Contactos.

Por un efecto de conflicto de interés el responsable del modelo dentro de la organización debe ser el área de riesgos, quien cumple funciones de custodio del modelo y además se encarga de validar la razonabilidad de los datos de cara a la dinámica del negocio. De igual forma esta área es también experta en el tema, y que, en conjunto con la mesa de dinero, conforman un equipo de especialistas que conocen a fondo la variable a explicar.

VII. Recursos de Entrenamiento.

Actualmente no se cuenta con equipos de computo especializados en el proceso de entrenamiento, por ello se deben usar los equipos de computo de uso diario por el área encargada del modelo, sin embargo, puede usarse cualquiera dado que se cuenta con el software Python el cual se puede usar indistintamente, sin ningún inconveniente.

Lo mencionado anteriormente, tiene el inconveniente de que, dada la restricción de máquinas estándar, no es posible cargar la totalidad de registros disponibles, sino que sólo es posible trabajar con un aproximado de 5 años de historia.

El presupuesto en términos monetarios es nulo, y en tiempo es limitado, lo cual puede hacer que se retrase la implementación del proceso y de igual manera el entrenamiento de este.

VIII. Despliegue.

El modelo se encuentra construido en lenguaje Python como se mencionó anteriormente, y con librerías de este, por tanto, tiene restricción para necesariamente tener este software instalado.

El resultado debe ser de uso diario, con lo cual su actualización se requiere necesariamente en las horas de la mañana de cada día hábil, esto no implica que el entrenamiento se realice diariamente, sino que la calificación de nuevas observaciones si se debe realizar día a día.

Los resultados deben estar disponibilidades para su consulta en un recurso compartido, y su visualización se puede realizar a través de algún proveedor de informes como Data Studio o Power BI.

En dado caso en el que se trabaje únicamente con los datos que se generan internamente en la organización, se debe evaluar la gestión de los datos que genera el sistema de valoración diaria de los fondos, y el cual contiene datos también de los niveles de retiros netos de los clientes.