

Pronóstico del retiro de un asociado a partir de las principales condiciones comerciales y socioeconómicas.

Forecast of the retirement of an associate based on the main commercial and socioeconomic conditions.

Néstor Bonilla Ramírez – nbonillar@libertadores.edu.co

Emerson Leonel Montero Vargas – elmonterov@libertadores.edu.co

José John Fredy González Veloza – jjgonzalezv02@libertadores.edu.co

RESUMEN

Para todas las cooperativas el número de asociados activos es la principal variable comercial de gestión y financiera a nivel patrimonial, ya que brinda una aproximación del potencial para colocar servicios de ahorro y crédito, además de ser el inicio de la intermediación financiera de la compañía.

En la actualidad, la cooperativa COONFIE presenta una baja retención de sus asociados. Identificando dos factores principales: el primero que está relacionado al rango de edad entre 50 a 70 años donde la retención disminuye debido a que es un segmento envejeciente que se compone principalmente por docentes que alcanzan la edad de pensión, generando un aumento creciente de la deserción o retiro. Y el segundo es la alta competencia bancaria para el mercado de créditos de empleados públicos que hace que la rotación de clientes aumente.

Entre los años 2016 y 2018, el porcentaje de retiros aumentó de 6.7% del total de asociados a 8.1%, un aumento de 1.4% en tres años. Los retiros pasaron de CO\$16,172 millones a CO\$17,869 millones entre 2016-2018, un aumento de CO\$1,697 millones que equivale a un 10%. Las anteriores cifras y la tendencia del aumento del porcentaje de retiros establecen la necesidad de tener una herramienta predictiva.

Mediante el modelo Random Forest (Geron A. 2019), se estimó los posibles retiros de asociados los cuales afectan el patrimonio de la cooperativa dado el creciente número de salidas de su base social.

Como resultado de este desarrollo estadístico se consiguió un modelo con un 69% de precisión y un 75% de exactitud para los asociados activos, igualmente se identificó que las variables que más afectan la retención de los clientes es el código de oficina, sexo, solicitud de certificado, edad y antigüedad, a la par se estructuró un nuevo esquema de proyección de pronósticos de retiro con los cuales el área comercial de la cooperativa trabajará en la retención de sus clientes, lo cual permitirá minimizar los esfuerzos comerciales ya que se realizará a grupos específicos identificados con este nuevo reporte predictor, que ayudará a reducir los costos operativos y aumentar la rentabilidad de la entidad financiera.

Palabras clave: Nivel patrimonial, Intermediación financiera, Segmento, Deserción, Random Forest, Precisión y Exactitud.

ABSTRACT

For all cooperatives, the number of active members is the main commercial management and financial variable at the equity level, since it provides an approximation of the potential to place savings and credit services, in addition to being the beginning of the financial intermediation of the company.

Currently, the COONFIE cooperative has a low retention of its members. Identified two main factors: the first is related to the age range between 50 to 70 years where retention decreases due to the fact that it is an aging segment that is mainly made up of teachers who reach pension age, generating a growing increase in desertion or withdrawal. And the second is the high banking competition for the public employee loan market that causes customer churn to increase.

Between 2016 and 2018, the percentage of withdrawals increased from 6.7% of total associates to 8.1%, an increase of 1.4% in three years. Withdrawals went from CO \$ 16,172 million to CO \$ 17,869 million between 2016-2018, an increase of CO \$ 1,697 million that is equivalent to 10%. The above figures and the trend of the increase in the percentage of withdrawals establish the need for a predictive tool.

Using the Random Forest model (Geron A. 2019), the possible retirements of associates were estimated, which affect the assets of the cooperative given the increasing number of exits from its social base.

As a result of this statistical development, a model was achieved with 69% precision and 75% accuracy for active associates, it was also identified that the variables that most affect customer retention are office code, sex, application of certificate, age and seniority, at the same time a new scheme of projection of retirement forecasts was structured with which the commercial area of the cooperative will work on the retention of its clients, which will allow minimizing commercial efforts since it will be carried out at specific groups identified with this new predictive report, which will help reduce operating costs and increase the profitability of the financial institution.

Keywords: Equity level, Financial intermediation, Segment, Attrition, Random Forest, Precision and Accuracy.

INTRODUCCIÓN

Para las cooperativas de ahorro y crédito de Colombia y el Mundo, el aporte social es el vínculo inicial que caracteriza a los potenciales clientes en asociados activos, los cuales forman la base social y patrimonial con la cual se fondea el 20% y 25% de la estructura financiera de estas entidades. (Confecoop, 2010).

Por cuenta de la alta competencia bancaria de los últimos 15 años se ha establecido un porcentaje de retiro o deserción ascendente de los asociados, los cuales son motivados a que no se tienen sistemas de relacionamiento o CRM que permita anticipar la identificación del retiro de ellos, así como establecer cuales ya cumplieron o están por cumplir la edad de pensión y que cambiaron de necesidades de servicio financiero.

Esto ha generado que la cooperativa COONFIE, presente un representativo retiro acumulado de asociados de más de 68.000 entre los años 2018 a 2020 y que ha marcado un alto cambio en el modelo de fondeo ya que es difícil sustituir el valor de los aportes de los asociados con más de 20

años de antigüedad versus los nuevos que se vinculan, ya que entran con un aporte mínimo de \$25.000 pesos.

Igualmente, se ha identificado que es mayor el costo operativo de vinculación de un nuevo afiliado que es aproximadamente \$215.000 pesos versus el costo de retención de un asociado antiguo que es de \$118.000 pesos.

Por lo anterior, este proyecto busca pronosticar la posibilidad de retiro de los asociados activos adultos de la cooperativa con el análisis de 4 variables operativas que inciden de forma directa y que fueron identificadas por el equipo comercial como lo es: El nivel de antigüedad, edad de pensión, solicitud de certificado de deuda, y que el saldo de aportes y ahorros sea igual o superior al valor de crédito.

Las cuales se integrarán a algoritmos de aprendizaje automático con énfasis de árboles de decisión con los cuales se podrá identificar la probabilidad de retiro por asociado y entregar esta información con periodicidad mensual o quincenal al equipo comercial de la cooperativa para que realice trabajo de fidelización y retención. (E.Alpaydin, 2010)

MARCO TEÓRICO

La normatividad, que existente nos canaliza a los principios, fundamentos y derecho consagrados en la Constitución Nacional de 1991, no obstante, el marco normativo de las cooperativas en Colombia posee una norma transversal que antecede la nueva carta fundamental, la Ley 79 de 1988 por medio de la cual se actualiza la legislación cooperativa y que aún a la fecha se mantiene activa y es la base legal del actuar del cooperativismo. (Coop – Cooperativas de las Americas, 2018).

La ley 454 de 1998 determina el marco conceptual que regula la economía solidaria, se transforma el Departamento Administrativo Nacional de Cooperativas en el Departamento Administrativo Nacional de la Economía Solidaria, se crea la Superintendencia de la Economía Solidaria, y el Fondo de Garantías para las Cooperativas Financieras de Ahorro y Crédito, se dictan normas sobre la actividad financiera de las entidades de naturaleza cooperativa y se expiden otras disposiciones.

Los asociados son el pilar de las cooperativas existentes, fundamentalmente, para satisfacer las necesidades de su base social activa conforme a los valores y principios que las definen como organizaciones cooperativas sin ánimo de lucro.

Las cooperativas financieras, cooperativas de ahorro y crédito y cooperativas multiactivas o integrales con sección de ahorro y crédito, se caracterizan por realizar actividad financiera, están obligadas a prestar el servicio de captación de dineros y colocación de créditos a sus asociados cumpliendo toda la normatividad en riesgos vigentes y facilitando todos los canales de atención necesarios para que disponga de sus recursos de forma segura.

Toda cooperativa debe tener unos aportes sociales mínimos no reducibles durante la vida de la cooperativa y estos deben fijarse en el estatuto. Es posible establecer contribuciones diversas según el tipo de cooperativa y la capacidad económica de los asociados. Los porcentajes y periodicidad de los aportes se establecerán dentro de los estatutos. Cuando se retira el asociado se entrega, en principio el total de lo aportado.

En los últimos años y ahora más por cuenta de la pandemia se está evidenciando un decrecimiento en los aportes sociales, lo cual afecta negativamente la dinámica de intermediación y debilita el patrimonio de estas entidades como se puede evidenciar en el siguiente gráfico: (Fecolfin,2021).

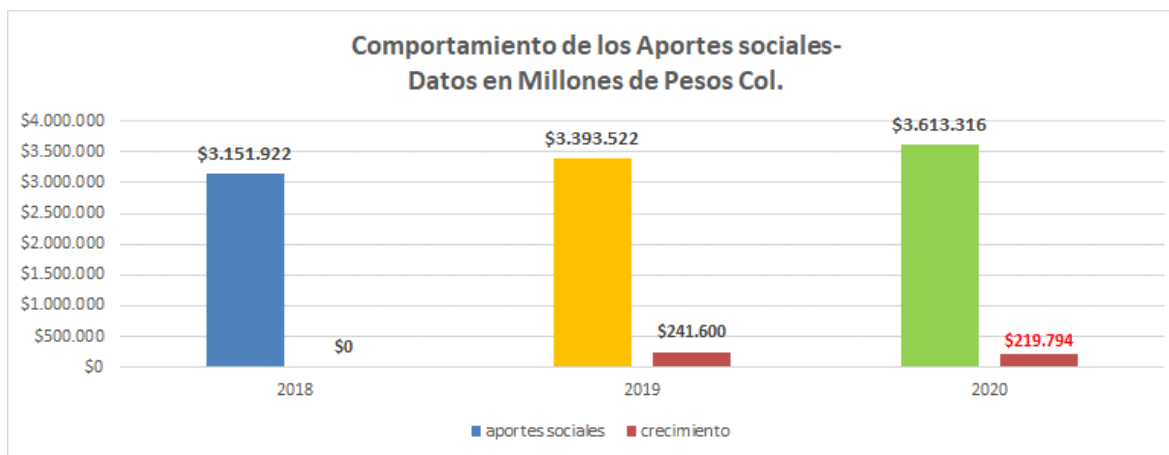


Figura 1. Comportamiento de aportes sociales cooperativas de Colombia: Elaboración propia.

De acuerdo con la figura 1 se observa que el crecimiento a diciembre de 2020 fue del 6,48%, frente a un 7,67% que fue el crecimiento porcentual en el 2019, lo cual muestra una reducción en su crecimiento del -1,19% equivalente a \$-21.806 millones que representa el impacto financiero del retiro creciente de los asociados.

- **Variables Socioeconómicas**

Las principales variables socioeconómicas y comerciales que influyen de forma directa o indirecta al momento del retiro de las cooperativas son las financieras, ya que muchos asociados están sobre endeudados y con la venta de su cartera a otras entidades bancarias buscan mejorar su flujo de efectivo personal alargando el plazo de las deudas vigentes aumentando su ingreso neto mensual, así mismo, los asociados que alcanzan la pensión y una edad de vejez alta, son reprimidos o aislados del servicio de crédito ya que muchas aseguradoras no cubren dichas solicitudes por cuenta de no poder suscribir el seguro de crédito ya sea por enfermedades preexistentes o comorbilidades las cuales restringe el servicio y solo le queda el ahorro personal, lo cual al final los motiva y obliga a tomar la decisión de retiro de las cooperativas.

METODOLOGÍA

Se definió trabajar con una metodología descriptiva cuantitativa utilizando Machine Learning como la herramienta indicada para potencializar la información de la cooperativa y aplicar por primera vez en la empresa un modelo automático predictivo para los retiros de asociados y aportes sociales, para lo cual se utilizará la herramienta Scikit-Learn (Scikitlearn.org, 2020).

Se realizaron modelos basados en árboles de decisión y que se describen a continuación:

Árboles de decisión:

Son algoritmos versátiles de aprendizaje automático que pueden realizar tanto, tareas de clasificación y regresión, e incluso tareas de salida múltiple. Donde se busca crear un modelo que prediga el valor de una variable objetivo con la aplicación de reglas de decisión simples que se establecen de las características de los datos. (L. Rokach y O. Maimond, 2015)

Se utiliza generalmente para clasificación de información, división y modelado de los posibles caminos tomados y su probabilidad de ocurrencia para mejorar su precisión. Una vez armados, los árboles de decisión se ejecutan muy rápido para traer los resultados.

Bosques Aleatorios (Random Forest):

Este modelo de aprendizaje está formado por un conjunto de árboles de decisión individuales, que son entrenados con una muestra ligeramente distinta de los datos de entrenamiento. La predicción de la variable de salida se obtiene agregando las predicciones de todos los árboles individuales que forman el modelo. (L. Rokach y O. Maimond, 2015)

Otra gran cualidad de los bosques aleatorios es que facilitan la medición de la importancia relativa de cada característica. Finalmente, Random Forest son muy útiles para comprender rápidamente qué características impacta la variable objetivo.

MODELO:

La aplicación del modelo planteado en este artículo se basó en la ejecución de 7 pasos de los 8 propuestos por (Zemsania 2018) para desarrollo de proyectos Machine Learning y que fue utilizado en el Dr. Carlos Huérfano Lenis, Estudiante de Maestría en Ingeniería y Analítica de Datos de la Universidad Jorge Tadeo Lozano: (Artículo: Identificación Predictiva de Fallos POS a través del Uso de Algoritmos de Aprendizaje Automático):

1. Construcción de la base y definición de variables.

El reporte y base de datos que se modeló se logró mediante conexión al ERP financiero propio de la empresa denominada (OPA- Ordenamientos Profesionales Automatizados), donde se consultó la información de 6 tablas (Asociados, Aportes o contribuciones, Captaciones, Cartera de Crédito, Certificados y Retiros) que contienen la mayor cantidad de registros (ver Anexo 1), esto obligó a realizar la construcción de una herramienta ETL de integración de datos de MICROSOFT llamada PowerBI, propiedad de la cooperativa para los procesos de extracción. (Amby.net, 2019)

Para el almacenamiento fue designado un espacio en la bodega de datos de la entidad la cual está ubicada localmente en los servidores de la cooperativa.

Una vez establecidos los campos de las tablas y su relacionamiento, como resultado de esta integración se generó un archivo con 38 variables y una etiqueta, para un total de 138.365 registros. Que contiene los asociados activos y retirados a nivel acumulado de los años 2018, 2019 y 2020 de la cooperativa incluido el saldo de los productos de aportes, depósitos y créditos vigentes y los certificados de deuda solicitados en el mismo periodo.



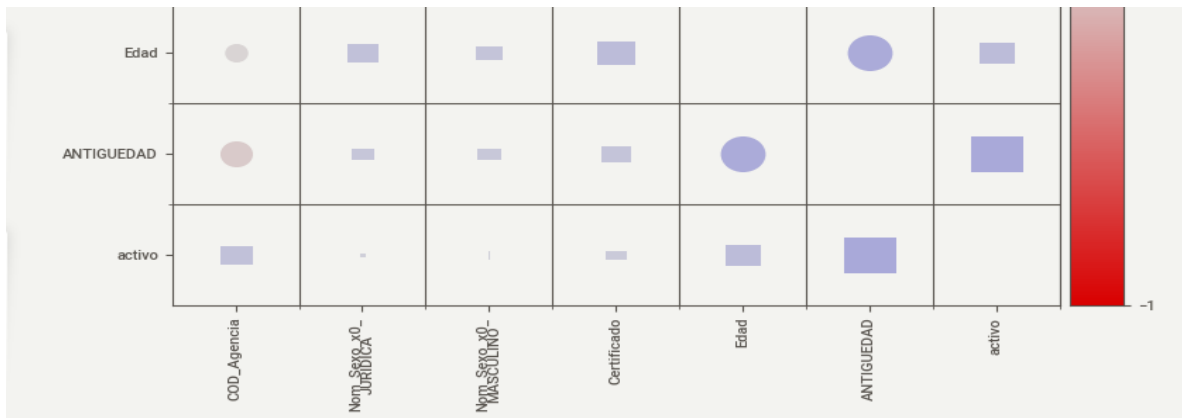


Figura 2. Gráfica de asociatividad de las variables de base de datos de la cooperativa. Fuente: Resultados obtenidos a través de Python librería Sweetviz

Como se evidencia en la Figura 2 la variable Activo tiene una relación fuerte con las variables Antigüedad, Edad y código de oficina, y una baja relación con las variables Certificado y Sexo, las cuales serán las utilizadas para los campos predictores del modelo.

3. Identificar correlaciones entre las variables independientes.

Con el análisis detallado y la identificación de la variable de salida ($y = \text{activo}$) se procedió a verificar las comparaciones del conjunto de datos y la integración completa de las correlaciones y asociaciones con las variables numéricas utilizando la correlación de Pearson y categóricas con el coeficiente de incertidumbre (ver Figura 3).

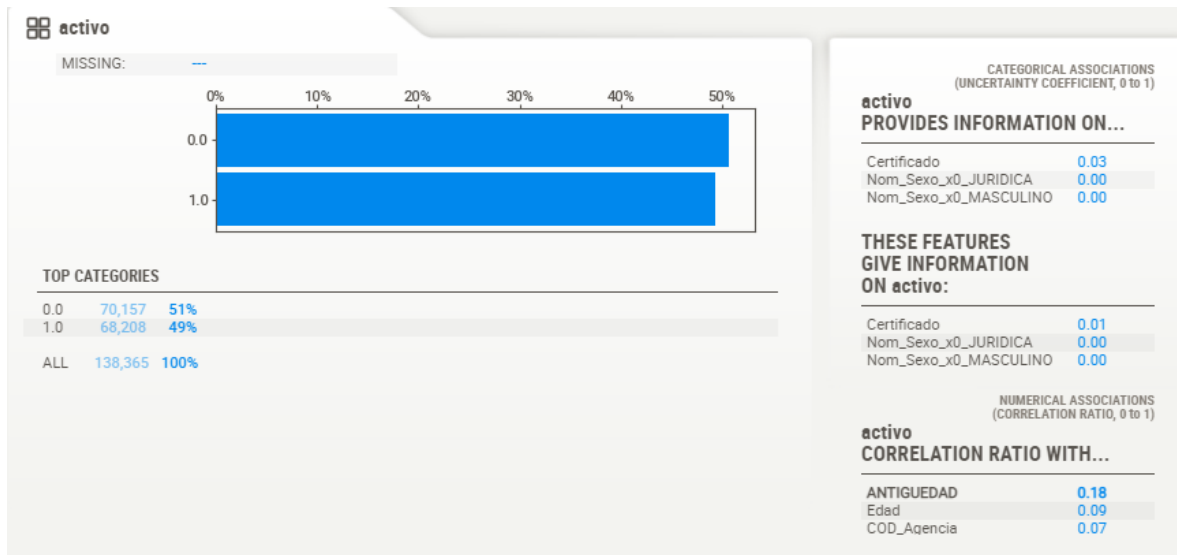


Figura 3. Gráfica de la variable “activo”. Fuente: Resultados obtenidos a través de Python librería Sweetviz

El proceso anterior generó una nueva transformación y ajuste en la base de datos ya que se depuró nuevamente y quedó con las siguientes variables las cuales se aplicaron para los modelos de algoritmos. (Anexo 4)

4. Aplicar el modelo de Machine Learning seleccionado

Con los datos listos e identificadas las características de la base y ratificado el modelo que se iba a trabajar, se dio paso a la construcción de este. Para esto se cargó la librería de Sklearn, y se procedió a realizar la formación del algoritmo, recordando que saldría correctamente si se elegían los parámetros adecuados y que garantizarían una buena calidad en los datos.

Las librerías que se ejecutaron fueron seleccionadas para la manipulación de la data y que se relaciona a continuación: NumPy y Pandas, para el manejo resultados que se iban dando como Matplotlib, la presentación de las métricas de Confusion matrix, roc_curve, auc, recall score, precision: score, f1_scores y la de los algoritmos seleccionados para la construcción de los modelos DecisionTreeClassifier y RandomForestClassifier (Anexo 5).

Previamente fue necesario transponer 6 variables categóricas para poder manejarlas como variables Dummy y que se relacionan a continuación: 'COD_Agencia', 'Nom_Sexo', 'Certificado', 'Edad', 'ANTIGUEDAD' y 'activo'. (Scikitlearn.org, 2020)

Seguidamente y con el ánimo de estandarizar las variables de entrada (X) del modelo se fijaron las columnas de predicción que se aplicarían en los comandos detallados (Anexo 6).

Igualmente se detalló la configuración de los parámetros de los clasificadores de los dos modelos, en el de árboles de decisión, la base de entrenamiento correspondía al 80% de la base y el de efectos de la prueba se tomó el 20%.

Los parámetros que se utilizaron en el modelo RandomForestClassifier fueron los siguientes (anexo 7):

N_estimators: Este parámetro hace referencia al número de árboles el cual se definió en 250.

Criterion: Se seleccionó Gini, porque generó un mayor porcentaje frente a entropy que mejora la sensibilidad.

Max_depth: Corresponde a la profundidad máxima de cada árbol el cual se dejó en 25 para nuestro caso.

Max_features: Número de predictores considerados en cada división, para nuestro caso se dejó None, que utiliza todos los predictores.

RESULTADOS

A continuación, se presentan los resultados en dos secciones: el primero referente al resultado de los modelos y su respectiva comparación y en el segundo se realizó el pronóstico para un grupo de asociados seleccionados por la entidad:

5. Comunicación y Evaluación del desempeño de los modelos seleccionados.

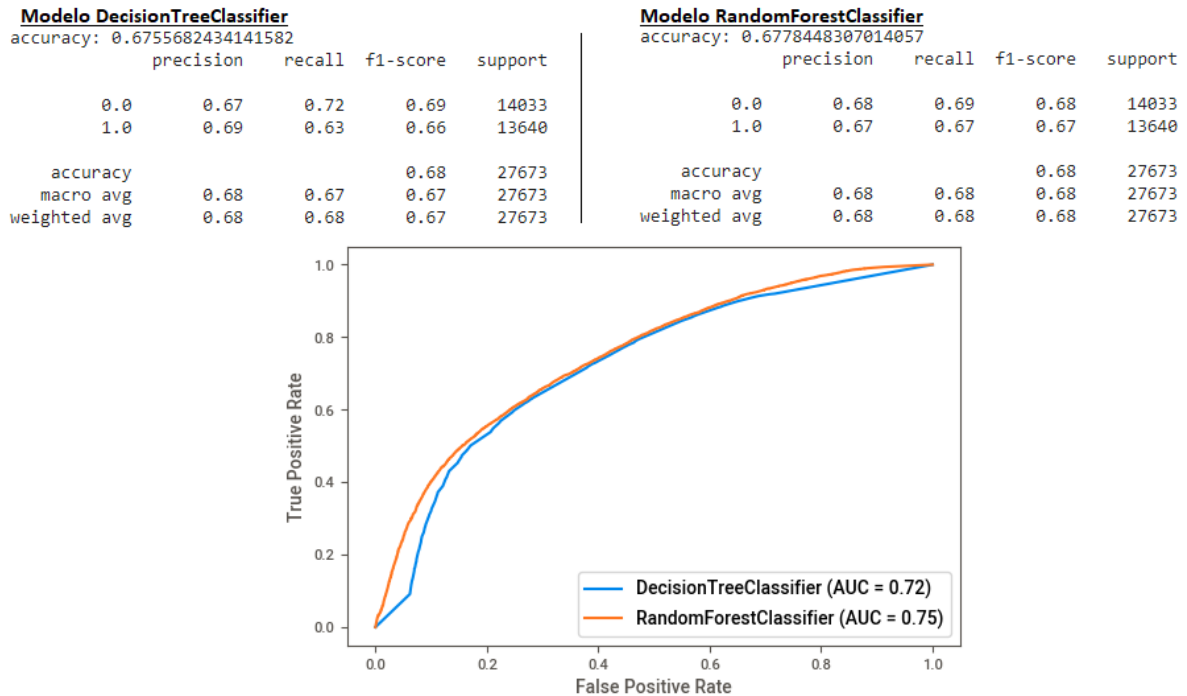


Figura 4. Curva ROC y AUC de los dos modelos DecisionTree y Random Forest. Fuente: Resultados obtenidos a través de Python

Al momento de evaluar los resultados de la figura 4 tomamos en cuenta dos indicadores F1-score que es el rendimiento combinado de la precisión y la exhaustividad y ACCURACY que es la exactitud del porcentaje de casos que el modelo ha acertado en la predicción, para nuestro caso el árbol de decisión tiene un mejor F1 = 0,72, sin embargo el ACCURACY es mayor en los árboles aleatorios 0,75 y dado que es la métrica más usada, se escogió este segundo modelo para llevar a la etapa de presentación ya que tiene el mayor rendimiento para ejecutar el modelo de pronóstico del posible retiro de asociados. (Jose Martinez Heras, 2020)

6. Ejecutar una prueba en conjunto con el área comercial.

Con el área comercial de la cooperativa se tomaron 12 casos de asociados para hacer una evaluación detallada a partir de los diferentes rangos de edad, antigüedad, código de oficinas y solicitud de certificados que se validaron contra el modelo de pronóstico Random Forest y se tuvieron los siguientes resultados, una vez ejecutado el comando predicción del modelo (Anexo 8):

caso	COD_Agencia	Certificado	Edad	ANTIGUEDAD	activo	% Proba. Retiro
1	1	NO	67	40	ACTIVO	0,50%
2	1	NO	61	38	ACTIVO	0%
3	2	NO	57	0	ACTIVO	89,63%
4	1	NO	56	36	ACTIVO	0%
5	3	SI	63	38	ACTIVO	0%
6	8	SI	26	2	ACTIVO	51,61%
7	4	SI	56	28	ACTIVO	4,40%
8	1	NO	65	19	ACTIVO	0%
9	1	NO	57	8	ACTIVO	48,19%
10	1	NO	23	6	ACTIVO	99,08%
11	1	NO	37	2	ACTIVO	47,31%
12	1	NO	39	12	ACTIVO	49,29%

Figura 5. Gráfico de tabla de resultados clientes de la entidad pronosticados con el modelo: Elaboración propia.

Se validaron cada uno de los casos pronosticados en función del porcentaje de probabilidad de retiro indicado en la figura 5 versus su estado actual (febrero 2021) con relación al corte de la base de datos de diciembre de 2020, detectando que, de los 12 casos, 10 están conforme a lo proyectado, solamente el caso 5 y el caso 9 no fueron acordes con lo proyectado, el quinto por muerte del asociado y el noveno por traslado de la oficina, lo cual indicaría que el porcentaje de exactitud del modelo es del 83,33%.

Igualmente se validó mediante encuesta telefónica varios de los asociados indicando cual sería la posibilidad a nivel porcentual de retiro de la cooperativa y se logró alcanzar mucha similitud en tres casos puntuales, indicando que si disminuía la tasa del crédito de consumo vigente se retirarían.

CONCLUSIONES

Se generó un reporte con información específica de los asociados que tienen la mayor probabilidad de salida por oficina, lo cual facilitará la atención focalizada y no general con los clientes, engranando con el plan de fidelización que tiene definido la subgerencia comercial, aumentando la retención de los asociados.

El solo hecho de saber cuál es el asociado a retener y comparar su costo versus el costo de vinculación de uno nuevo generaría un ahorro de \$97.000 por cliente, que, si se multiplica por el número de retiros del año 2020 que fue de 6.287, se tendría un ahorro a nivel de los gastos comerciales de \$610.000.000 aproximadamente.

La administración de la cooperativa debe garantizar que los funcionarios comerciales asimilen esta nueva actividad y tomen esta información predictiva para que se genere un impacto positivo en los indicadores y reduzca el índice de retiro o deserción de asociados a un porcentaje menor del 5% del total de asociados activos.

Por cuenta de la influencia de la variable certificado, antigüedad y edad en la probabilidad de retiro se estableció un reporte interno en la cooperativa con el cual se está generando un trabajo comercial previo, donde se trabaja comercialmente en la fidelización para atenuar los falsos positivos y falsos negativos del modelo que inicia su uso y que tendrá seguimiento en un periodo de seis meses para validar su impacto y correspondiente ajuste.

Finalmente debemos recordar que todo proyecto tiene que mantener el mejoramiento continuo como premisa, es bueno el resultado inicial de lo alcanzado por el modelo creado, pero se debe verificar el mejoramiento de la calidad de la información de la cooperativa ya que se tuvo que

realizar limpieza y transformación de variables que dificultan la estandarización de los datos y que podían sesgar los cálculos.

RECOMENDACIONES

Todo modelo de machine learning requiere que los funcionarios encargados de generar los pronósticos tengan conocimiento amplio de las herramientas tecnológicas, por lo anterior es importante seguir capacitando ya que se debe mantener en los colaboradores una visión innovadora que permita dimensionar el potencial de los datos que tiene la cooperativa, y aprovechar este nuevo recurso comercial pensado en los clientes y ampliarlo a nuevos procesos como el de riesgos de crédito con la generación de un modelo propio de motor de decisión en el otorgamiento de crédito con el apoyo de las diferentes técnicas estadísticas.

AGRADECIMIENTOS

Se agradece al área de tecnología y transformación digital su gran compromiso en la estructuración de los recursos necesarios para la generación de los reportes que usamos en el pronóstico del retiro de los asociados.

REFERENCIAS BIBLIOGRÁFICAS

- Geron A. (2019) Aurelien Geron - Chapter 7. Ensemble Learning and Random Forests - Hands-on Machine Learning with Scikit-Learn, Keras, and TensorFlow (2019, O'reilly) - libgen.lc
- E. Alpaydin. (2010) Introduction to Machine Learning second edition. Massachusetts Institute of Technology.
- Confecoop. (Enero de 2010). *Las Cooperativas de Ahorro y Credito en Colombia*. Obtenido de Observatorio cooperativo No. 14. Recuperado de: <http://www.confecoop.coop/observatorio/14/files/doc14.pdf>

Coop – Cooperativas de las americas. (Diciembre de 2018). *Informes de Marcos Legales de Colombia*. Obtenido de confecoop. Recuperado de: <https://confecoop.coop/wp-content/uploads/2020/03/Informe-de-Marcos-Legales-Colombia.pdf>

Fecolfin. (Diciembre de 2020). *Cooperativas de Ahorro y Credito los Resultados en el Año de la Pandemia*. Investigación No. 28. Recuperado de: https://mcusercontent.com/caa1f6d4de782a0ba4bc049bd/files/2f0faae7-ed61-4fac-a313-b845f4d83548/RESULTADOS_TOTAL_COOPERATIVAS_DE_AHORRO_Y_CRE_DITO_A_DICIEMBRE_DE_2020_2_1_.pdf

Scikitlearn.org. (2020) Scikit-learn 0.24.1 Recuperado de: http://scikitlearn.org/stable/supervised_learning.html#supervised-learning.

L. Rokach y O. Maimond. (2015) *Data Mining With Decision Trees Theory and Applications*. World Scientific. Segunda Edición. 2015

Zemsania. (2018) *Pasos para desarrollar un proyecto de Machine Learning*. Recuperado de: https://www.zemsania.com/recursos-zemsania/whitepapers/DTS/Machine_learning.pdf. 2018. Referencia #24 De Artículo (Identificación Predictiva de Fallos POS A Través del Uso de Algoritmos de Aprendizaje Automático) De: Carlos Huérfano Lenis. Maestría en Ingeniería y Analítica de Datos. Universidad Jorge Tadeo Lozano.

Amby.net. (2019) *Power BI – Herramienta de Business Intelligence*. Recuperado de: <https://amby.net/2019/02/18/power-bi-herramienta-de-business-intelligence/>

Pypi.org. (2021) sweetviz 2.1.0. Recuperado de: <https://pypi.org/project/sweetviz/>

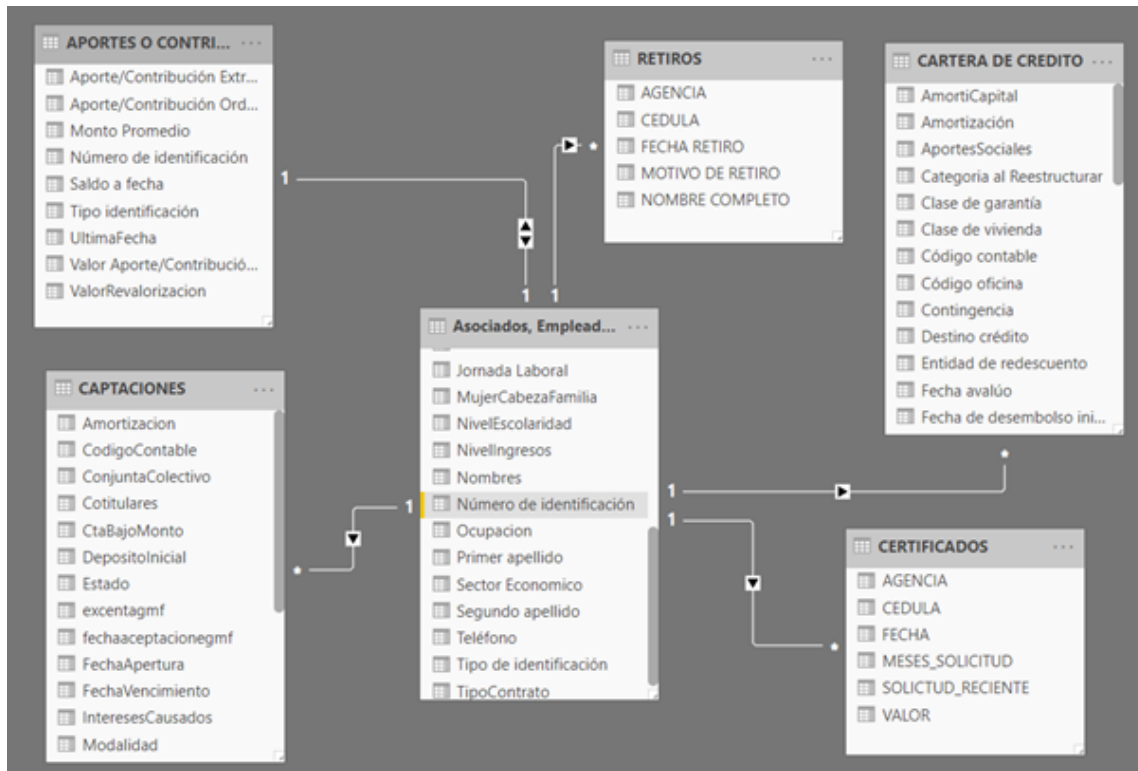
Scikitlearn.org. (2020) Scikit-learn 0.24.1 Recuperado de: <https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.OneHotEncoder.html>



Jose Martinez Heras. (2020) Precision, Recall, F1, Accuracy en clasificación. Recuperado de: <https://www.iartificial.net/precision-recall-f1-accuracy-en-clasificacion/>

ANEXOS:

- 1- Modelo de bases de datos relacional para la construcción del reporte para el pronóstico del retiro de asociados: Elaboración propia.



- 2- Ajuste de datos y nuevas variables de la tabla base de datos. Fuente: Resultados obtenidos a través de Python.

```
df['patrimonio'] = df['saldoTotalAportesLinea_I'] + df['SaldoTotalAhorros'] - df['SaldoCapitalTotalCreditos']

df['Telefono1'] = df.Telefono1.astype('category')
df['Celular'] = df.Celular.astype('category')
df['Nombre_TipoVivienda'] = df.Nombre_TipoVivienda.astype('category')
df['Nroacta'] = df.Nroacta.astype('category')
```

- 3- Tipos de variables. Fuente: Resultados obtenidos a través de Python.

#	Column	Non-Null	Count	Dtype
---	-----	-----	-----	-----
0	COD_Agencia	138365	non-null	int64
1	NOM_Agencia	138365	non-null	object
2	NUMERO_IDENTIFICACION	138365	non-null	int64
3	Nom_TipoIdentificacion	138365	non-null	object
4	FechaExpcedula	129706	non-null	object



5	NombreIntegrado	138365	non-null	object
6	Nom_Sexo	138365	non-null	object
7	FechaNacimiento	138365	non-null	object
8	FechaNacimiento.1	138365	non-null	object
9	Edad	138365	non-null	int64
10	Telefono1	138365	non-null	category
11	Celular	138365	non-null	category
12	Direccion	138365	non-null	object
13	Estrato	138229	non-null	float64
14	Nombre_TipoVivienda	138365	non-null	category
15	Nom_UbicacionVivienda	100353	non-null	object
16	Cod_Ciiu	138365	non-null	int64
17	Nombre_Ciiu	126514	non-null	object
18	Nombre_Segmento	138269	non-null	object
19	Nombre_EmpresaNomina	112244	non-null	object
20	Nombre_Estudios	137135	non-null	object
21	Nombre_Profesion	108893	non-null	object
22	Nombre_EstadoCivil	109622	non-null	object
23	Nroacta	138365	non-null	category
24	FechaActa	138365	non-null	object
25	ANTIGUEDAD	138365	non-null	int64
26	Nombre_EmpresaLabora	109105	non-null	object
27	fechaingresoLabora	92645	non-null	object
28	Nombre_TipoContrato	109336	non-null	object
29	FechaTerminaContrato	34178	non-null	object
30	Nombre_Cargo	108910	non-null	object
31	NOM_TIPO_PERSONA	138365	non-null	object
32	NOM_LINEA	137858	non-null	object
33	Certificado	138365	non-null	object
34	activo	138365	non-null	object
35	saldoTotalAportesLinea_T	138365	non-null	float64
36	SaldoTotalAhorros	138365	non-null	float64
37	SaldoCapitalTotalCreditos	138365	non-null	float64
38	patrimonio	138365	non-null	float64

- 4- Librerías cargadas para la generación del modelo. Fuente: Resultados obtenidos a través de Python

```
[ ] import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from sklearn.model_selection import train_test_split
from sklearn.metrics import confusion_matrix
from sklearn.metrics import classification_report
from sklearn.tree import DecisionTreeClassifier
from sklearn.ensemble import RandomForestClassifier
```

- 5- Variables de entrada utilizadas para los modelos de algoritmos.

#	Column	Non-Null	Count	Dtype
---	-----	-----	-----	-----
0	COD_Agencia	138365	non-null	float64
1	Nom_Sexo_x0_JURIDICA	138365	non-null	float64
2	Nom_Sexo_x0_MASCULINO	138365	non-null	float64



3	Certificado	138365	non-null	float64
4	Edad	138365	non-null	float64
5	ANTIGUEDAD	138365	non-null	float64
6	activo	138365	non-null	float64

6- Variables predictoras de los modelos.

```
columns_predic=['COD_Agencia','Nom_Sexo_x0_MASCULINO','Certificado','Edad','ANTIGUEDAD']
```

7- Configuración final del modelo Random Forest. Fuente: Resultados obtenidos a través de Python.

```
from sklearn.ensemble import RandomForestClassifier
modelo_rf = RandomForestClassifier(
    n_estimators=250,
    criterion='gini',
    max_features= None,
    max_depth = 25 )
```

8- Comando de predicción del modelo de árboles aleatorios.

```
modelo_rf.predict_proba(X_prospec_trasnform[['COD_Agencia','Nom_Sexo_x0_MASCULINO','Certificado','Edad','ANTIGUEDAD']])
```