MODELO PREDICTIVO PARA DETERMINAR LOS RETIROS DE UNIFORMADOS EN EL GRADO DE PATRULLERO DE LA POLICÍA NACIONAL DE COLOMBIA

RUBÉN ALEXIS PORTILLO ROSERO, **raportillor@libertadores.edu**

ANDREA MONTAÑA RAMÍREZ, **amontanar@libertadores.edu**

JULIÁN DAVID LÓPEZ PIÑEROS, **jdlopezp01@libertadores.edu**

RESUMEN

La presente investigación busca identificar un modelo predictivo, que permita representar la realidad de los retiros de uniformados en el grado de patrullero de la Policía Nacional de Colombia, a partir de la identificación de patrones, según características sociodemográficas y condiciones laborales individuales de este personal, con el fin de predecir futuros comportamientos de retiro mediante el análisis machine learning. Se analizaron datos 2011 a 2021, desde un enfoque cuantitativo, siguiendo un diseño "no experimental" transversal de alcance analítico-descriptivo, a partir de la evaluación de métricas de los modelos Árbol de Decisión, Regresión Logística, Soporte de Maquina Vectorial -SVM, K-Nearest Neighbors-KNN y Choosing Features, catalogándose este último como el modelo predictivo, con mejor resultado en sus métricas, destacando una exactitud del 85%, insumo que sirve para potenciar las estrategias de fidelización y retención del talento humano policial.

Palabras Claves: machine learning, métricas, modelos predictivos, Patrulleros Policía Nacional, retiros.

ABSTRACT

The present investigation seeks to identify a predictive model, which allows representing the reality of the withdrawals of uniformed officers in the patrol degree of the National Police of Colombia, based on the identification of patterns, according to sociodemographic characteristics and individual working conditions of this personnel. in order to predict future withdrawal behaviors using machine learning analytics. Data from 2011 to 2021 were analyzed from a quantitative approach, following a cross-sectional "non-experimental" design with an analytical-descriptive scope, based on the evaluation of metrics from the Decision Tree models, Logistic Regression, Support Vector Machine -SVM, K-Nearest Neighbors-KNN and Choosing Features, the latter being classified as the predictive model, with the best result in its metrics, highlighting an accuracy of 85%, an input that serves to enhance the strategies of loyalty and retention of police human talent.
Keywords: machine learning, metrics, predictive models, National Police Patrollers, withdrawals.

INTRODUCCIÓN

Uno de los intangibles más importantes para las empresas en general, es su Talento Humano, para la Policía Nacional, es el capital que más se atesora, es por ello que a partir de la presente investigación se busca predecir escenarios futuros, frente al retiro anticipado de los patrulleros.

Para tal fin se hace uso del machine learning, el cual brinda nuevas formas de aprovechar la inteligencia artificial, para mejorar las organizaciones, sus procesos y hacer que las actividades de sus colaboradores sean mucho más sencillas, con el fin de ahorrar tiempo y dinero. Es por ello que en la modernización, cobran importancia la automatización, mejorar los desempeños, estandarizar los procedimientos y aprovechar la información para la toma de decisiones, a partir de modelos matemáticos, cuya finalidad es que los algoritmos que incorporen, sean capaces de realizar tareas inteligentes.

Por lo anterior es fundamental realizar un pronóstico que permita determinar las características del personal de uniformados de la Policía Nacional en el grado de Patrullero, que se retiran por solicitud propia de manera anticipada, basado en las características sociodemográficas de este personal, por lo que

resulta de vital importancia determinar y medir en cierto grado la incertidumbre asociada, la intención de retiro, logrando evidenciar que características presenta una mayor probabilidad de retiro, lo que conlleva a adentrarse en el mundo de la modelación empleando técnicas de vanguardia en el análisis de datos.

Un modelo de predicción realizado con técnicas computacionales, que ayude a inferir las diferentes situaciones previas a su presentación, permitirá seleccionar el mejor modelo que pronostique si un funcionario tiene tendencia a retirarse de la Institución y así derivar propuestas frente a las estrategias de retención y fidelización a implementar, para no perder el capital humano y contribuir en el cumplimiento de la misionalidad institucional en materia de convivencia y seguridad ciudadana.

REFERENTES TEÓRICOS

Se aborda el tema de modelo predictivo aplicable a la fidelización de personal desde dos ópticas: 1ª. Deserción laboral, en la cual se busca tener un contexto, parámetros y tipos de modelos empleados que permitan enriquecer esta investigación y sirvan como insumo argumental y 2ª. Desde el campo policial en el cual converge toda la información que se logre obtener, teniendo como referente la deserción anticipada en cuerpos de policía y los factores que inciden en ello.

Desde años atrás se ha contemplado el estudio de la deserción y como aplicar modelos para predecir su comportamiento, en Estados Unidos, la investigación *"Predicting Attrition: An Empirical Study at the United States Air Force Academy*" describe un marco analítico a través del cual se puede predecir la deserción voluntaria, el enfoque incorpora la Teoría Marshalliana de la Utilidad y un procedimiento de Estimación de Máxima Verosimilitud, para evaluar la propensión al desgaste de un individuo específico (Dempsey, 1976).

Aunado a lo anterior, la investigación denominada "*Un modelo para predecir la retención y el desgaste de los hispanoamericanos en la Marina"* desarrolló un modelo longitudinal para explicar y predecir las decisiones de permanecer o salir de la Marina, abordando las decisiones desde el punto de

vista del individuo, concluyendo que quedarse o irse depende de la interacción continua de variables de fondo, sentimientos al ingresar, experiencias en la Marina y un sentido general de compromiso con la Marina. (Temple Univ Philadelphia Pa Merit Center, 1980).

En Oriente Medio, para Mohamed (2022), en su investigación sobre la "*gestión del conocimiento organizacional y recursos humanos preparados para el futuro: identificación de predictores de jubilación anticipada en una fuerza policial moderna*", orientada a un cuerpo de policía del continente asiatico, a través del empleo de datos secundarios (encuesta), se realizó la evaluación de componentes demográficos, nivel de satisfacción con el trabajo, y las circunstancias sociales, empleando un modelo de regresión logística, logró identificar el desequilibrio entre el trabajo y la familia, destacando criterios como el número de hijos menores de 18 años, el desequilibrio entre el trabajo y la vida privada, y la disminución de la frecuencia de encuentros con amigos, están asociados con la intención de jubilación anticipada.

En México, se realizó un análisis sobre "*La deserción laboral de los millennials en México: un nuevo paradigma en las organizaciones"* utiliza el método cualitativo, de carácter descriptivo, cuyo tipo de investigación es correlacional, para identificar cuáles son las mayores necesidades y satisfacciones de la generación estudiada (Garza, 2019).

(Terreros, Vega, & Pupo, 2019) en su estudio "*Aplicación del Teorema de Bayes en la selección de personal para disminuir la deserción laboral"* aplicaron una encuesta a una muestra de 384 personas, que se procesó a través del rechazo de casos, conteo de frecuencias y cálculo de probabilidades, en el cual se tuvieron en cuenta variables como sexo, edad, escolaridad, problemas lumbares, tipo de desvinculación laboral, concluyendo que la red bayesiana de deserción laboral es una herramienta de soporte del proceso de selección de personal.

(Reyes, 2019) proponen en su trabajo de investigación "*Predicción de deserción laboral utilizando un algoritmo genético y redes neuronales artificiales"* una herramienta que permita identificar la posible deserción de un empleado. El método propuesto consiste en un algoritmo genético que identifica las variables relevantes y mejora la arquitectura de una red neuronal artificial como solución, los cuales permitieron predecir la deserción laboral hasta un 88,92 % de exactitud, utilizando variables como: salario, ambiente laboral, línea de carrera, satisfacción laboral, beneficios, compromiso organizacional, relación con superiores, estrés laboral, estado civil y género.

Según el informe "sobre el equilibrio que alcanzan los ciudadanos entre el trabajo y su vida personal", se realizó un estudio a 20 países acerca de la importancia del equilibrio entre la vida personal y el trabajo, citado por Aristizábal (2019) y realizado por la Organización para la Cooperación y el Desarrollo Económico (OCDE), el resultado arrojó en una escala de 0 a 10 que Colombia tiene un 0.9% de equilibrio entre la vida personal y su trabajo, verificando variables de compromiso familiar, número de horas laborales, pocas horas de ocio, satisfacción personal y una casi nula flexibilización laboral; estos factores desencadena una baja productividad, deserciones, cierre de empresas, índices de pobreza en ascenso, trabajo informal (freelance), entre otros factores (Ducange, 2016).

Desde la línea de investigación de la psicología organizacional, la publicación denominada *"Deserción de oficiales y suboficiales con cargos operativos que pertenecen a la escuela militar de aviación Marco fidel Suarez de la fuerza Aérea Colombiana",* con enfoque mixto tipo descriptivo, por medio de la aplicación de instrumento tipo encuesta, se concluyó que pensamientos relacionados con los horarios, turnos laborales establecidos por los superiores, y la relación de los beneficios establecidos por la entidad, son los generadores de la deserción laboral. (Pinto, Malagon, Castro, Elizalde, & Redondo, 2020)

Por otra parte, frente a las investigaciones relacionadas con la temática a tratar en la Policía Nacional encontramos un "Estudio exploratorio sobre el retiro a solicitud propia de agentes de la Policía

Nacional por desmotivación" las cuales abordan la temática desde un enfoque analítico descriptivo. (Linares, Ramírez, & Camelo, 1984)

Adicional a lo anterior, encontramos que en la Escuela de Estudios Superiores de Policía (ESPOL), se realizó la investigación denominada *"Diagnóstico al proceso de retiro de personal en la Policía Nacional para plantear estrategias que reduzcan los reintegros en la institución"*, con un enfoque analítico, mixto de corte descriptivo, en el cual se establecen los diferentes casos en los cuales los retiros se convierten en futuros reintegros. (Bolaños, Franco, & Marin, 2009)

La investigación *"Perfil del retiro laboral temprano en población uniformada de la Policía de Colombia (2008-2018)"* es utilizado un método cuantitativo de tipo corte transversal analítico aplicando el modelo de regresión logística, abordado con un enfoque de salud como base explicativa del retiro, en el cual se integran variables como edad, sexo, estado civil, categoría policial, grado policial, área laboral, unidad laboral, año y mes de retiro. Como variable dependiente se tomó el tipo de retiro, el cual se clasificó en cinco (5) tipos: 1. retiro voluntario y por calificar servicios, 2. retiro por discapacidad e incapacidad médica, 3. retiro por destitución o incumplimiento, 4. retiro por incapacidad académica, y 5. retiro por causas inesperadas, estudio en el cual concluye que el retiro laboral temprano es el más frecuente en la institución. (Ceballos & Campos, 2020)

En la publicación "*Análisis de los factores que inciden en la deserción laboral de pilotos y tecnólogos pertenecientes al área de aviación policial*" producto de la investigación basada en enfoque mixto, se utilizó un instrumento de tipo encuesta, se detectaron causas económicas y familiares como las que ocasionan la deserción laboral. (Pimienta, Carvajal, & Calderon, 2016)

Para concluir, el aprendizaje autónomo propuesto puede aplicar a numerosas investigaciones. Las aplicaciones a problemas supervisados y no supervisados se pueden encontrar abundantemente en la literatura.

Los métodos para la predicción de lo propuesto en la presente investigación están constituidos fundamentalmente en la aplicación de cinco métodos de machine learning, los mismos comprenden una serie de datos históricos y un análisis univariado, así como multivariado.

Para el presente documento es preciso tener en cuenta las siguientes palabras las cuales hacen parte del contexto institucional, así:

De acuerdo a lo definido en el Decreto No. 1791 14/09/2000 "Por el cual se modifican las normas de carrera del Personal de Oficiales, Nivel Ejecutivo, Suboficiales y Agentes de la Policía Nacional", en el cual se establecen catorce causales de retiro las cuales se describen como: por solicitud propia, por llamamiento a calificar servicios, por disminución de la capacidad sicofísica, por incapacidad absoluta y permanente o gran invalidez, por destitución, por voluntad del Gobierno para oficiales y del Ministro de Defensa Nacional o la Dirección General de la Policía Nacional por delegación, para el nivel ejecutivo, los suboficiales y los agentes, por no superar la escala de medición del Decreto de Evaluación del Desempeño Policial, por incapacidad académica, por desaparecimiento, por muerte, por no superar la validación de competencias, por decisión judicial o administrativa, por inhabilidad y por separación absoluta.

Adicional a lo anterior se encuentra en el contexto institucional palabras propias de la dinámica de la institución, como lo son:

- **Vínculo familiar:** sin vinculo corresponde a solteros, separados, divorciados y viudos; los de vínculo familiar son funcionarios casados y unión libre.
- **Nivel misional:** se presenta los niveles en los cuales se desempeñan los policías, que se describen:
  - Operativo: se considera en este nivel, al personal que por la naturaleza del cargo que desempeña desarrolla procesos encaminados a contrarrestar el accionar de la delincuencia común y

organizada, responsable directo de la prevención y conservación del orden público o convivencia pacífica, en misiones de vigilancia, Inteligencia, Policía Judicial, Servicio Aéreo, labores primarias del servicio policial, lucha contra el Secuestro, Extorsión y Narcotráfico, Policía Comunitaria, Tránsito y Medio Ambiente.

- Administrativo: se considera al personal que desarrolla tareas relacionadas con la administración de los recursos, de talento humano o de servicio y apoyo.
- Docente: se considera al personal que desarrolle actividades de formación, capacitación y especialización en la Escuela Nacional de Policía "General Santander", y sus Seccionales y colegios de Bienestar Social de la Policía Nacional.

- **Aptitud:** Es el conjunto de habilidades, destrezas, aptitudes y potencialidades de orden físico y psicológico, que deben reunir los uniformados para ingresar y permanecer en el servicio (Decreto 1796 del 14/09/2000, articulo No. 3)

  - Apto: se refiere a la persona que presente condiciones psicofísicas que permitan desarrollar normal y eficientemente la actividad militar, policial o civil, correspondiente a su cargo, empleo o funciones.
  - No Apto: son las condiciones del uniformado, frente a si presenta disminuciones de la capacidad psicofísica que le impida laborar en unidades operativas y se reubique en cargos administrativos.

- **Sigla física:** Corresponde a la unidad donde el uniformado labora físicamente sin importar su vinculación administrativa a una especialidad de policía.

- **Retiro:** Es la situación por la cual el personal uniformado, sin perder el grado, cesa en la obligación de prestar servicio.

- **Grado**: también llamado rango militar, graduación o empleo, es un sistema jerárquico para establecer la escala de mando de las fuerzas armadas, fuerzas policiales y otras organizaciones armadas o uniformadas.

- **Pie de fuerza:** hace referencia a la cantidad de personas sobre la que se forma un cuerpo militar.

METODOLOGÍA

La presente investigación es de tipo "cuantitativo", se desarrolla a partir del análisis machine learning de los datos de retiros de patrulleros de la Policía Nacional de Colombia durante la década 2011-2021, sigue un diseño "no experimental" en la cual no hay intervención directa del investigador, "transversal" al recopilar información en un tiempo determinado, de alcance "analítico-descriptivo" al agregar datos históricos, a partir de la evaluación de métricas de los modelos Árbol de Decisión, Regresión Logística, Soporte de Maquina Vectorial-SVM, K-Nearest Neighbors-KNN y Choosing Features.

Por consiguiente, la metodología se desarrolló en cuatro fases, así: 1ª. Presentación de la base de datos y descripción de las variables. 2ª. Análisis exploratorio de los datos y componentes principales. 3ª. Diseño de la herramienta de pronóstico. 4ª. Implementación del modelo que permita inferir que funcionarios de policía son susceptibles al retiro por solicitud propia en el grado de Patrullero.

Fase 1. Presentación de los datos

Se realizó la validación en bases de datos institucionales, para verificar cuales de las variables tratadas en los anteriores estudios pueden ser de acceso para el presente estudio y que aporten en la implementación de estrategias para la fidelización del personal. Es preciso indicar que se generaron varias

solicitudes internas entre dependencias, con el fin de poder agrupar la mayor cantidad de información, ya que institucionalmente se cuenta con diferentes administradores de bases de datos y no se encuentran relacionadas entre sí.

Referente a los datos es importante resaltar que son de carácter privado y no está autorizada su divulgación por tratarse de información de los funcionarios y para su divulgación se requiere autorización expresa de los particulares y de la institución.

Para el estudio se empleó software libre en R - Studio Cloud y Python Software Foundation License versión 3.10.5, con el fin de llevar a cabo el análisis y las gráficas, así como la visualización de diferentes códigos de machine learning para el modelado de los datos.

Criterios de inclusión y Exclusión

Se analizó la información de los patrulleros de la Policía Nacional de Colombia, que ingresaron a la institución entre el año 2011 y 2021, obteniendo un total de 66.780 uniformados, de los cuales 4.036 pidieron su retiro por solicitud propia y 2.982 fueron solicitudes por las diferentes causas. Información extraída del Sistema para la Información del talento Humano (SIATH) de la Policía Nacional.

Inclusión: base de datos policiales, con información del personal en el grado de Patrullero, en los cuales se tuvo en cuenta fecha de ingreso a la institución periodo comprendido entre los años 2011 a 2021 y que se encontraban reportados en la base de datos SIATH (Sistema de Información para la Administración del Talento Humano) Policía Nacional de Colombia, así mismo, su situación laboral (activo – retirado) al momento de su extracción.

Exclusión: policías retirados por las diferentes causales excepto (solicitud propia), como el personal reintegrado a la institución o con otras situaciones administrativas del servicio.

Variables: se analizarán variables como: sexo, ascenso, año de nacimiento, lugar de nacimiento, región donde labora, lugar de trabajo vs lugar de origen, zona de orden público, tiempo de servicio, tiempo en la última unidad, cantidad de unidades laboradas, nivel misional de trabajo, grupo étnico, estado civil, nivel académico, hijos, aptitud, edad, felicitaciones, condecoraciones, sanciones, embargos, clima laboral entre otros. Por otra parte, se toma como variable respuesta "RETIRO".

Adicional a los anterior es importante mencionar la distribución de los Departamentos de la geografía colombiana por Regiones de Policía, en las cuales encontramos ocho regiones de Policía y para el presente estudio se toma como unidades independientes el Departamento de Amazonas y el Archipiélago de San Andrés y Providencia.

Tabla 1

Distribución por regiones de Policía

UNIDADES POLICIALES

REGIÓN	DEPARTAMENTO
REGIÓN 1	Boyacá Cundinamarca
REGIÓN 2	Tolima Huila Caquetá Putumayo
REGIÓN 3	Caldas Quindío Risaralda
REGIÓN 4	Valle del Cauca Nariño Cauca
REGIÓN 5	Norte de Santander Santander Arauca
REGIÓN 6	Antioquia Choco Córdoba Urabá
REGIÓN 7	Meta

	Casanare Guainía Guaviare Vichada
REGIÓN 8	Atlántico Bolívar Guajira Magdalena Cesar Sucre
AMAZONAS	Amazonas
SAN ANDRES Y PROVIDENCIA	San Andrés

Nota: Las Regiones de Policía y los Departamentos no corresponden necesariamente a la División Político Administrativa de la Nación. Amazonas y San Andrés y Providencia por organización están integradas a la Región de Policía 1, sin embargo, para este estudio se presentan como independientes.

Fase 2. Análisis Exploratorio de los datos

En el siguiente análisis se evidencian los diferentes métodos estadísticos que se desarrollaron a lo largo de la investigación, en su etapa preliminar del procesamiento de los datos, así: inicialmente se observa el análisis descriptivo para verificar el comportamiento de las variables, seguido del análisis univariado en el cual se describe la distribución de cada variable y sus métricas, por último se realizó el análisis multivariado con el fin de verificar la dimensionalidad y la relación de dependencia de las variables.

A continuación, se presenta el consolidado de la última década (2011-2021), de los retiros por todas las causales, una vez analizada la información se tomó como insumo para el estudio la causal de retiro servicio activo por solicitud propia, como aparece en la tabla siguiente:

Tabla 2

Causales de Retiro

CAUSALES DE RETIRO (2011-2021)

CAUSAL	TOTAL	%
Solicitud propia	4.036	57,5%
Voluntad de la dirección general	1.073	15,3%
Destitución	789	11,2%
Muerte en servicio activo	608	8,7%
Disminución de capacidad psicofísica	239	3,4%
Incapacidad absoluta permanente, gran invalidez	121	1,7%
Otras inhabilidades ley 734 de 2002	117	1,7%
Separación absoluta	33	0,5%
Desaparecimiento	1	0,01%
Llamamiento a calificar servicios	1	0,01%
Total	**7.018**	**100%**

En la Tabla No.2, se observa que 4.036 policías se han retirados de manera anticipada por solicitud propia en la última década correspondiente al (57.5%), del total de retiros presentados por todas las causales (7.018). Por otra parte, los retiros por solicitud propia corresponden al (6%) del total de personal que ingreso a la institución (66.780), de los cuales en este momento se encuentran en servicio activo (59.762) uniformados en el grado de patrulleros.

Respecto al comportamiento anual de los 4.036 (57.5%) retiros por solicitud propia y tiempo de servicio, relacionado a la variable tiempo de servicio se observa el siguiente impacto:

Tabla 3

Retiros por Solicitud propia según tiempo de servicio

COMPORTAMIENTO RETIROS POR SOLICITUD PROPIA Y TIEMPO DE SERVICIO						
INGRESO	**CANTIDAD**	**ACTIVOS**	**RETIROS**	**AÑOS DE SERVICIO**	**% RETIRADO**	**% AÑO**
2011	7881	7245	636	12	8,8%	0,7%
2012	8488	7782	706	11	9,1%	0,8%
2013	8458	7756	702	10	9,1%	0,9%
2014	10321	9319	1002	9	10,8%	1,2%
2015	3197	2967	230	8	7,8%	1,0%
2016	4483	4171	312	7	7,5%	1,1%
2017	4259	4039	220	6	5,4%	0,9%
2018	3649	3558	91	5	2,6%	0,5%
2019	4341	4272	69	4	1,6%	0,4%
2020	6673	6609	64	3	1,0%	0,3%
2021	2050	2046	4	2	0,2%	0,1%
total	**63800**	**59764**	**4036**	**-**	**6,8%**	**0,7%**

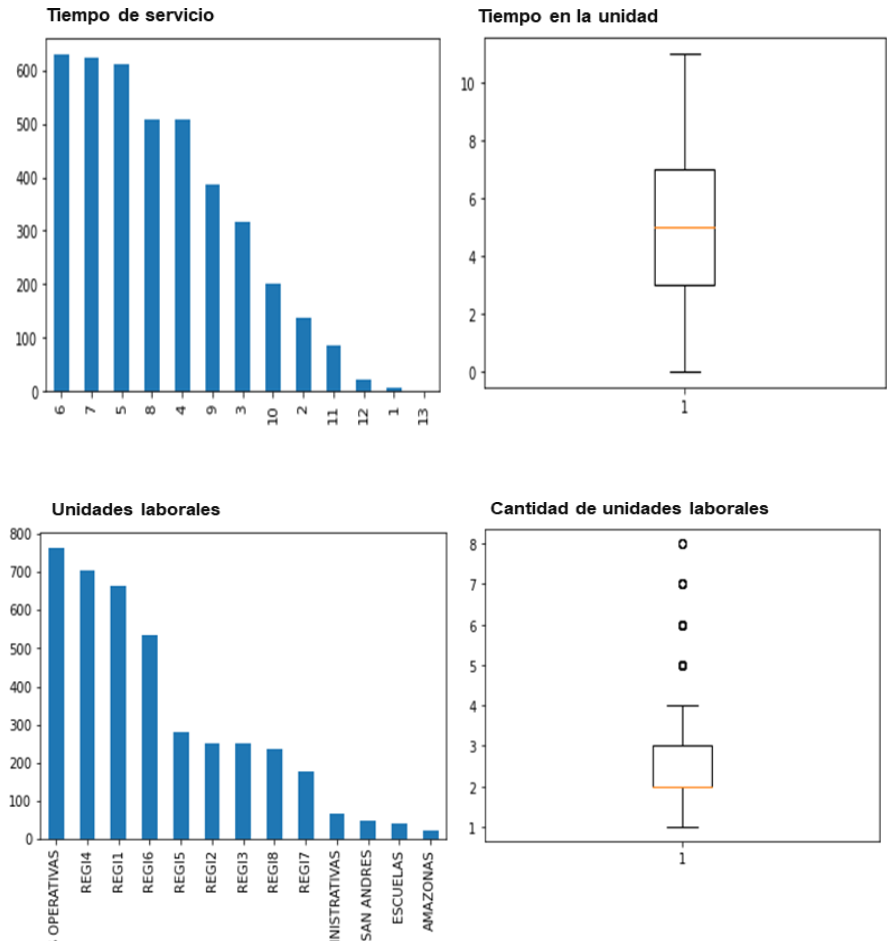
La tabla No. 3, presenta los porcentajes de retiro por tiempo de servicio, en los cuales se puede inferir que en promedio se retira anualmente en 0.7% del personal que ingresa por cada año, así mismo hay más retiros de uniformados en el periodo de seis a diez años de servicio.

A continuación, se realiza una descripción de variables sociodemográficas del personal, arrojaron como resultado de manera descriptiva, datos que ilustran la toma de decisiones en materia de priorización de las poblaciones a impactar con estrategias de fidelización y retención de talento humano.

De las variables de la situación laboral entre las que se destacan: unidades laborales, tiempo de servicio, tiempo en la unidad y cantidad de unidades laboradas, como aparece a continuación:

Figura 1

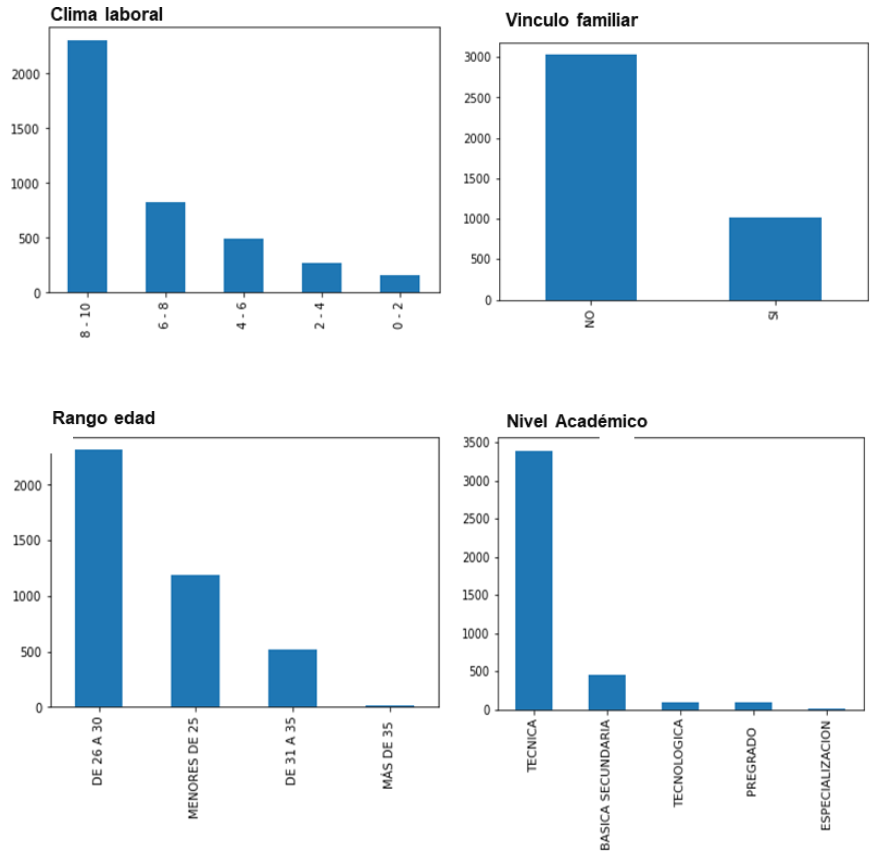
Resultados de las variables de las condiciones laborales del personal retirado



Como resultado del análisis de datos del personal retirado por solicitud propia, se identificó ciertas características de esta población frente a las condiciones laborales, evidenciando que el 71% contaban con un tiempo de servicio entre 4 y 8 años, siendo el periodo de tiempo más propenso para el retiro; frente a las unidades laborales más afectadas están las Direcciones Operativas, seguido de la Región Cuatro y la Uno que en conjunto representan el 53%, con relación al tiempo en la última unidad encontramos al mayor concentración en el periodo de tres a siete años el 63% y prestaban su servicio en unidades de orden público, el 74% pertenecía al nivel operativo, el 51% no trabajaba cerca al lugar de nacimiento y por último en el diagrama de cantidad de unidades laboradas la mayor concentración de datos se encuentra de dos a tres equivalente al 86% y datos atípicos como 5, 6, 7 y 8 unidades.

Figura 2

Resultados de las variables sociodemográficas



Frente a las variables sociodemográficas de la población objeto de estudio, se evidencia el 87% corresponde a hombres y el 13% a mujeres; el 75% registran sin vínculo familiar (soltero, divorciado, separado y viudo), seguido del 25% con vínculo familiar (casados y unión libre); el 57% se encuentran en el rango de edad entre 26 a 30 años; el 84% contaban con un nivel académico de técnico; el 98% eran aptos para el servicio; el 39% tenían hijos y frente al clima laboral el 57% se encuentra en el rango de 80% a 100% de satisfacción.

Con la utilización del Power BI Desktop, versión 2.105.1143.0, se eléboro un mapa de calor como una técnica de visualización de datos para presentar la magnitud del impacto por unidades de policía a causa de los retiros a solicitud propia. La variación del color presentada por su intensidad, arrojo los siguientes resultados:

Figura 3

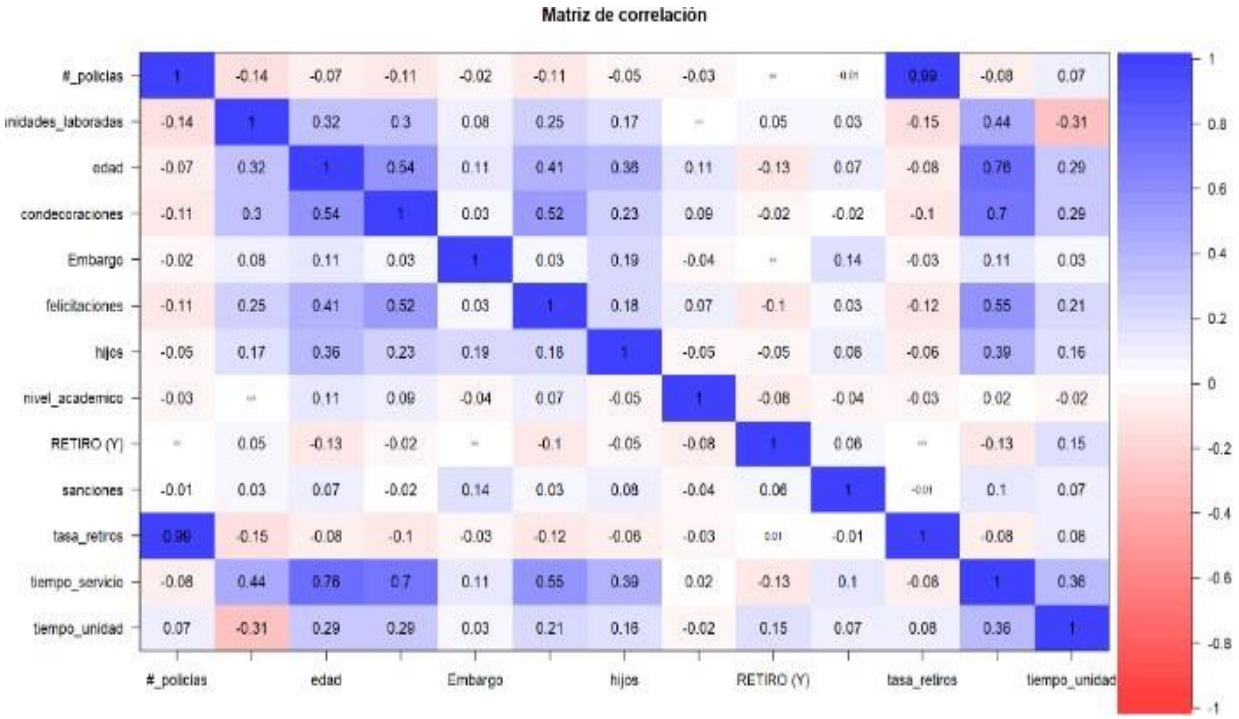
Índice de retiro construido en Power BI Desktop (Mapa de Calor)

Se observan los 32 departamentos geográficos que corresponden a 61 unidades de policía, donde los más afectados por el índice de retiros son: Cundinamarca con (0.059), Valle del Cauca (0.055), Risaralda (0.055) Antioquia (0.054) y el archipiélago de San Andrés y Providencia con (0.55), así mismo, para las unidades administrativas se evidencia que la afectación más significativa se presenta en la Oficina de Telemática (0.075), seguido de la Dirección de Protección y Servicios Especiales (0.074). Es de anotar que la media del índice de retiros por unidad en promedio es de (0.010)

Los anteriores datos muestran que predomina el retiro anticipado en los lugares de más complejidad delincuencial, en los cuales se tiene un alto riesgo para laborar como uniformado, predomina el sexo masculino al ser por tasa el más representativo, la mayoría de ese personal se encontraban sin vínculo familiar vigente y con un tiempo de servicio entre 4 a 8 años.

Figura 3

Matriz de Correlación – (R. Studio)



Se observa una correlación de variables entre las que se destacan, tasa de retiros con unidades (0.99), edad con tiempo de servicio (0.77), felicitaciones con tiempo de servicio (0.59), condecoraciones con tiempo de servicio (0.72), tiempo de servicio con unidades laboradas (0.47), edad con condecoraciones (0.57).

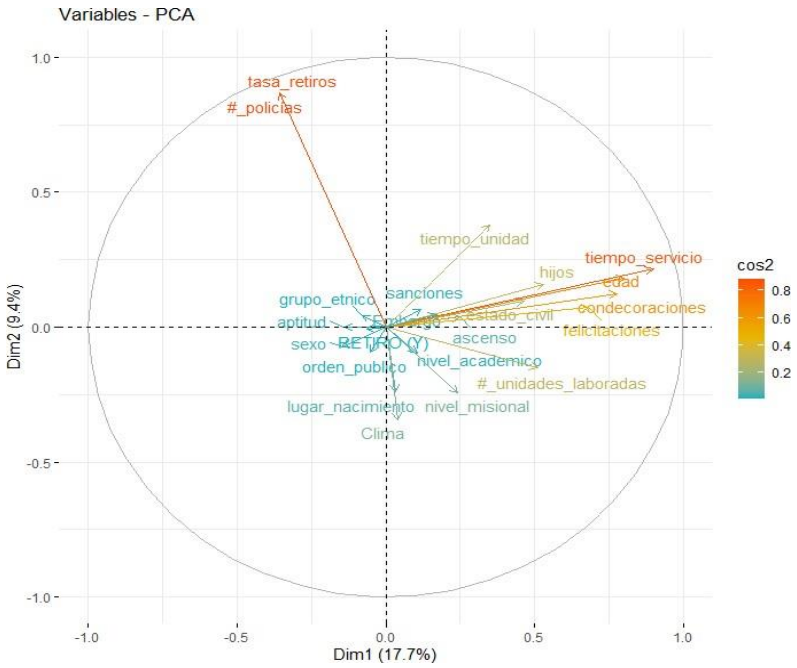
Continuando; con el desarrollo del trabajo investigativo se aplica la técnica para reducir la dimensionalidad del conjunto de datos propuesto, para que la misma intuitivamente sirva para hallar las causas de la variabilidad del conjunto de datos y ordenarlos por importancia, de la siguiente manera:

Análisis de Componentes principales

Se realiza la codificación de las variables categóricas, con el fin de realizar el análisis de componentes principales y verificar la dimensionalidad de los datos.

Figura 4

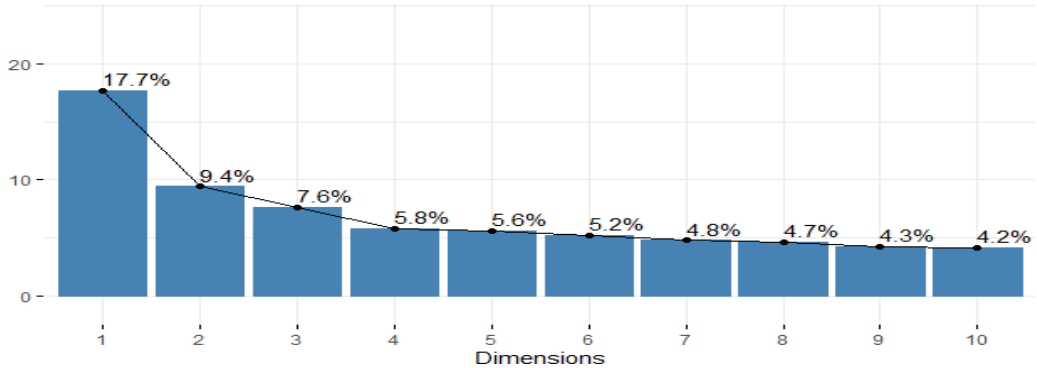
Dimensionalidad de variables. – (R. Studio)



Las variables son poco relacionadas y en las dos primeras dimensiones representan tan solo en el 27.1%, lo que complica explicar los datos con pocas dimensiones ya que se pierde información.

Figura 4

Histograma dimensionalidad de variables. – (R. Studio)



Los datos se representan en un total de 22 dimensiones, sin embargo, en la gráfica se presentan las primeras diez primeras, donde la dimensión uno se encuentra variables como tiempo de servicio, edad,

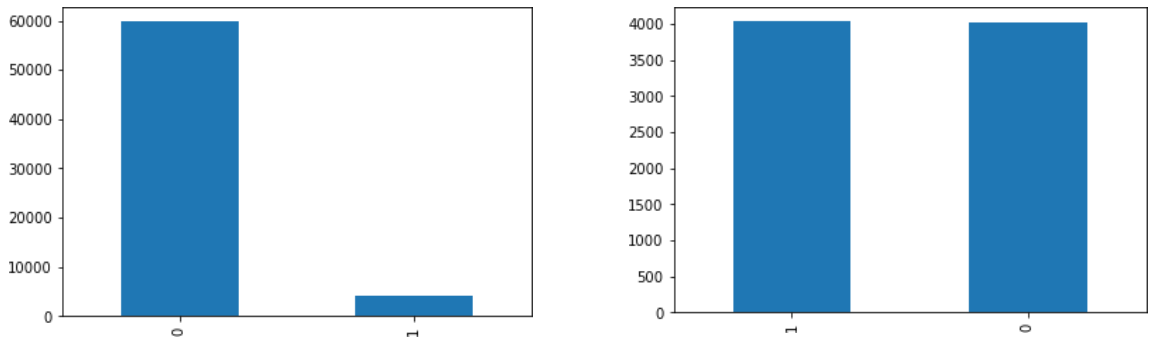
condecoraciones, felicitaciones, hijos, unidades laboradas, estado civil, unidad, tiempo en la unidad, así mismo, en la segunda dimensión encontramos tasa de retiros, unidad, tiempo en la unidad, clima laboral, nivel misional, lugar de nacimiento, tiempo de servicio, edad, hijos, unidades laboradas.

Fase 3. Diseño de Herramienta para el Pronóstico

Para la creación del modelo pronostico se requirió aplicar técnicas de desbalanceo, toda vez que al contener un 93.6% de datos en servicio activo y tan solo el 6.4% de datos retirados (variable respuesta), se hizo necesario emparejar los datos aplicando un submuestreo.

Figura 5

Desbalanceo de los datos.



La gráfica del costado izquierdo muestra el desbalanceo de los datos, el cual causa que la predicción se incline hacia los datos de mayor porcentaje, para lo cual se aplican metodologías de submuestreo estratificado se logra extraer diferentes muestras iguales para relacionar con las variables respuesta y evitar el desbalanceo, quedando la base 50 – 50 con relación a la variable respuesta.

Fase 4. Implementación de los Modelos

Para llevar a cabo la implementación de los modelos, es importante afirmar que el aprendizaje automático es una rama en evolución de los algoritmos computacionales que están diseñados para emular la inteligencia humana aprendiendo del entorno circundante. Las técnicas basadas en el aprendizaje automático se han aplicado con éxito en diversos campos que van desde el reconocimiento de patrones, la

visión por computadora (El Naqa, I., Murphy, MJ 2015), con ello se busca identificar patrones en los uniformados que permitan mediante el análisis poder resolver el problema de investigación propuesto en este proyecto. Para lo cual se plantea el procesamiento de los datos a partir de los modelos supervisados.

En este sentido, el Aprendizaje supervisado, son algoritmos que trabajan con datos "etiquetados" (labeled data), intentando encontrar una función que, dadas las variables de entrada (input data), les asigne la etiqueta de salida adecuada. El algoritmo se entrena con un conjunto de datos etiquetados y así "aprende" a asignar la etiqueta de salida adecuada a un nuevo valor, es decir, predice el valor de salida. El aprendizaje supervisado suele emplearse en problemas de clasificación y regresión.

Por lo tanto, en esta fase se procedió a formular cada uno de los modelos con la información que surtió del tratamiento del desbalanceo, teniendo como referentes técnicos el entrenamiento y validación (70%-30%), así como, sus métricas (accuracy o recall), matriz de confusión y comparación de los mismos a través de método ROC:

Árbol de decisión: Es una técnica de aprendizaje inductivo supervisado no paramétrico, se utiliza para la predicción y se emplea en el campo de inteligencia artificial, donde a partir de una base de datos se construyen diagramas de construcción lógica, muy similares a los sistemas de predicción basados en reglas, que sirven para representar y categorizar una serie de condiciones que ocurren en forma repetitiva para la solución de un problema (Giudici, Paolo. 2003).

XGBoost Extreme Gradient Boosting: es un algoritmo predictivo supervisado que utiliza el principio de boosting, por lo que la importancia detrás de este principio es crear múltiples modelos de predicción "débiles" secuencialmente, tomando los resultados del modelo anterior, y de esta manera generar un modelo más "fuerte", logrando el mejor poder predictivo y mayor estabilidad en sus resultados (Bosco, 2020).

Máquinas de vectores de soporte (Support Vector Machines): la aplicación de este tipo de modelos se ha extendido a problemas de clasificación múltiple y regresión, destacando que el SVMs ha resultado ser uno mejor clasificador para un amplio abanico de situaciones, considerándose uno de los referidos dentro del machine learning y su base se sustenta en el concepto de hiperplano (Rodrigo, 2020).

K-Nearest Neighbor (KNN): es uno de los métodos más simples para resolver problemas de clasificación, de aprendizaje supervisado no paramétrico, en la fase de aprendizaje almacena las tuplas de entrenamiento, sin realizar ningún cálculo y espera hasta que se le dé una tupla de prueba para clasificar. (Adeniyi, Wei y Yongguan 2016).

Modelo de regresión logística: es un instrumento estadístico de análisis multivariado, de uso tanto explicativo como predictivo. Resulta útil su empleo cuando se tiene una variable dependiente dicotómica (un atributo cuya ausencia o presencia hemos puntuado con los valores cero y uno, respectivamente) y un conjunto de variables predictoras o independientes, que pueden ser cuantitativas (que se denominan covariables o covariadas) o categóricas. (Chitarroni, 2002)

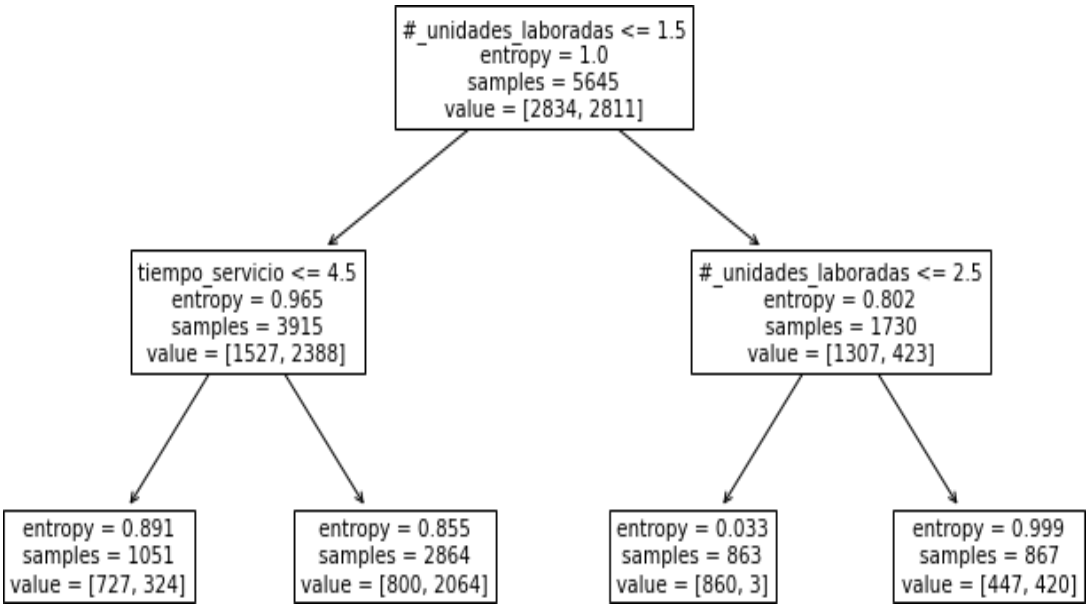
Evaluación de métricas, aparte de verificar la importancia de cada modelo, permite comparar distintos modelos con el fin de elegir el mejor, por tanto, probar si un clasificador es bueno, es revisando el porcentaje de acierto en la clasificación de los datos. En este sentido la Matriz de confusión muestra el desempeño de un clasificador describiendo como se distribuyen los valores reales y las predicciones. Otro aspecto que cabe destacar es el empleo de la curva ROC, que a través de un gráfico se muestra el rendimiento de un modelo de clasificación, relacionando la sensibilidad con el porcentaje de falsos positivos (Abella, 2021).

RESULTADOS

Figura 6

Árbol de decisión.

```
|--- #_unidades_laboradas <= 1.50
|   |--- tiempo_servicio <= 4.50
|   |   |--- class: 0
|   |--- tiempo_servicio >  4.50
|   |   |--- class: 1
|--- #_unidades_laboradas >  1.50
|   |--- #_unidades_laboradas <= 2.50
|   |   |--- class: 0
|   |--- #_unidades_laboradas >  2.50
|   |   |--- class: 0
```

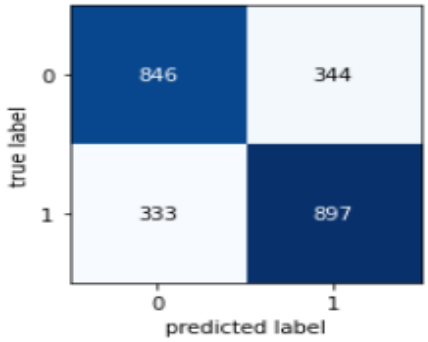


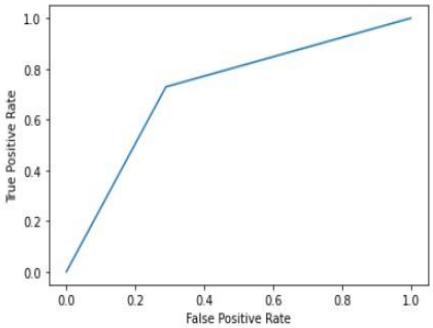
En la gráfica se observa un nodo raíz con 5645 muestras, que hace una primera subdivisión a la izquierda por número de unidades laboradas que sea menor a 1.5 y a la derecha mayores a 1.5, en el segundo nivel a la izquierda el tiempo de servicio menor a 4.5 con 3.904 muestras, a la derecha la cantidad de unidades laboradas con 1.741 muestras.

Figura 7

Árbol de decisión matriz de confusión

	precision	recall	f1-score	support
0	0.7176	0.7109	0.7142	1190
1	0.7228	0.7293	0.7260	1230
accuracy			0.7202	2420
macro avg	0.7202	0.7201	0.7201	2420
weighted avg	0.7202	0.7202	0.7202	2420



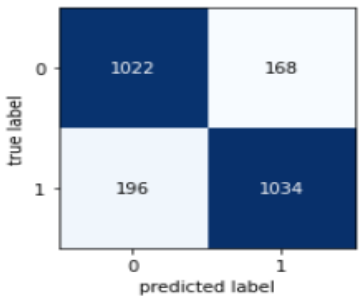


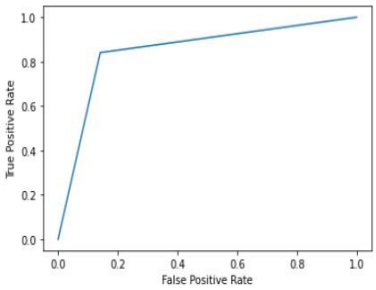
Se evalúan las diferentes métricas expuestas anteriormente en las cuales el modelo está prediciendo con una precisión del 72,28% para los retiros, con una sensibilidad de 71,09% en los activos y 72,93 para los retiros y un a exactitud del 72,02%. La matriz clasifica 897 retiros reales, y 846 activos reales, sin embargo, se tiene 344 falsos positivos y 333 falsos negativos.

Figura 8

XGBoost Extreme Gradient Boosting matriz de confusión – curva ROC

	precision	recall	f1-score	support
0	0.8391	0.8588	0.8488	1190
1	0.8602	0.8407	0.8503	1230
accuracy			0.8496	2420
macro avg	0.8497	0.8497	0.8496	2420
weighted avg	0.8498	0.8496	0.8496	2420



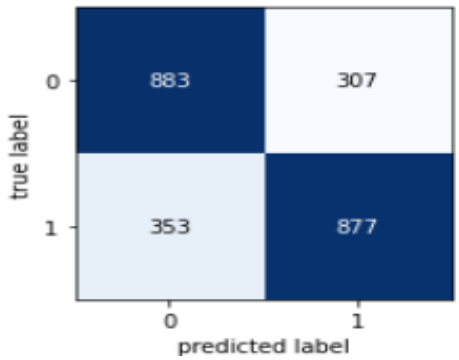


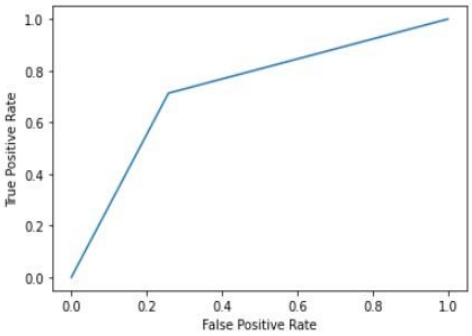
Emplear este tipo de modelo de clasificación, como se muestra en la Figura 9, se logró una precisión del 86,02% para los retiros, con una sensibilidad de 84,88% en los activos y 85,03% para los retiros y un a exactitud del 84,96%. La matriz de confusión clasifica 1034 retiros reales, y 1022 activos reales, pero se evidencia 168 falsos positivos y 196 falsos negativos.

Figura 9

Máquinas de vectores de soporte (Support Vector Machines) matriz de confusión – curva ROC

	precision	recall	f1-score	support
0	0.7144	0.7420	0.7279	1190
1	0.7407	0.7130	0.7266	1230
accuracy			0.7273	2420
macro avg	0.7276	0.7275	0.7273	2420
weighted avg	0.7278	0.7273	0.7273	2420





La exactitud que se evidencia en la Figura 9, en este tipo de modelo es de 72,73%, clasificando como retiros reales a 877 funcionarios y 883 activos reales, no obstante, 353 son catalogados como falsos negativos.

Figura 10

K-Nearest Neighbor (KNN) matriz de confusión – curva ROC

	precision	recall	f1-score	support
0	0.7334	0.7815	0.7567	1190
1	0.7743	0.7252	0.7490	1230
accuracy			0.7529	2420
macro avg	0.7539	0.7534	0.7528	2420
weighted avg	0.7542	0.7529	0.7528	2420

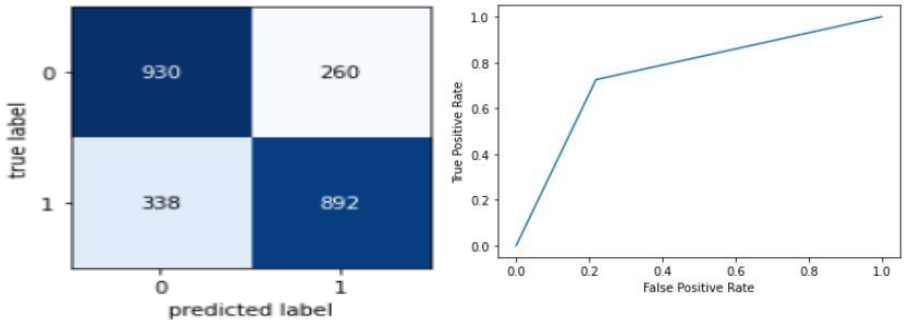
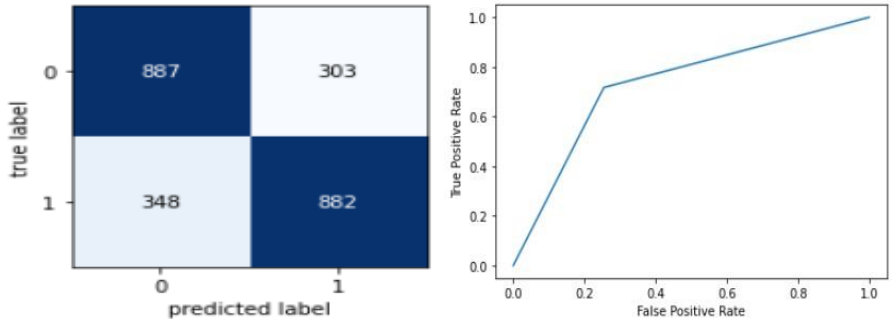


Figura 11

Modelo de regresión logística matriz de confusión – curva ROC

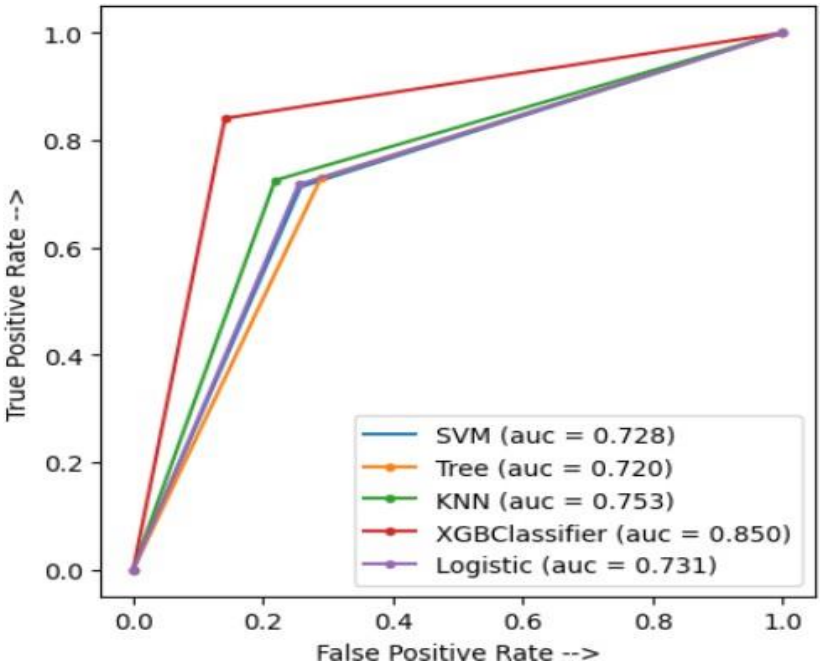
	precision	recall	f1-score	support
0	0.7182	0.7454	0.7315	1190
1	0.7443	0.7171	0.7304	1230
accuracy			0.7310	2420
macro avg	0.7313	0.7312	0.7310	2420
weighted avg	0.7315	0.7310	0.7310	2420



En las figuras 10 y 11 se muestran los resultados de aplicar los modelos de referencia, encontrando que los niveles de exactitud están en el 75.29 y el 73.10% respectivamente, además los valores de precisión están alrededor del 71 al 78 % para estos modelos.

Figura 12

Curva ROC – comparación de Modelos



Después de analizar cada modelo por separado, establecer sus porcentajes y aspectos de relevancia, se unificaron las métricas de cada uno de ellos y bajo el contexto de este grafico de la curva, en la Figura 12, se destaca el modelo **XGBoost Extreme Gradient Boosting**, por tener los mejores valores en comparación con el resto de ellos, por lo que en este estudio se puede compartir lo afirmado por Bosco (2020) en lo concerniente al poder predictivo y mayor estabilidad que otros modelos de clasificación.

DISCUSIONES

De acuerdo con la metodología establecida, se puede determinar que es viable realizar investigaciones aplicando modelos para predecir escenarios futuros, toda vez que, aplicados los modelos propuestos con el enfoque prospectivo, donde los datos de entrada y salida corresponden a un mismo instante en el tiempo (Maguire, Batty et al., 2005) (datos o variables propuestos de los funcionarios de policía) y los modelos dinámicos, donde la salida corresponde a una respuesta que pueda indicar la injerencia de las variables en la toma de decisiones del policía para solicitar el retiro, siendo los modelos predictivos utilizados, redes neuronales, arboles de decisión y regresión logística.

Así mismo, se considera viable la estructuración del modelo a partir de las variables indicadas, atendiendo lo señalado en el estudio titulado *Predicción de deserción laboral utilizando algoritmos genéticos y redes neuronales artificiales* (Reyes, 2019), en el cual se realiza el comparativo de las variables más utilizadas en diferentes estudios de deserción laboral, evidenciando que han sido analizadas variables como salario, ambiente laboral, línea de carrera, satisfacción laboral, beneficios, compromiso organizacional, relaciones con los superiores, stress laboral, estado civil y género.

Según Fidalgo y Gouveia (2011, citado en Reyes, 2019), existen dos tipos de variables que contribuyen a la deserción laboral: aquellas que las empresas pueden controlar, tales como salario, horario, línea de carrera, etc.; y las que no puede controlar como reubicación, problemas personales, entre otros. En concordancia con lo anterior, para el presente modelo se tienen en cuenta variables similares las cuales basado en el análisis de expertos se ha evidenciado que afectan directamente el proyecto de vida del empleado y su plan de carrera institucional.

Aunque la metodología utilizada demanda mayor limpieza de datos y procesos de transformación de variables para ser ejecutadas computacionalmente, se logra el objetivo de aplicar los modelos propuestos.

Los resultados obtenidos con relación al comportamiento de la predicción de los retiros del personal de la Policía Nacional en el grado de Patrullero estudiados mediante modelos predictivos, son claros y abren la puerta para futuras investigaciones de las demás causales de retiro y otras poblaciones en diferentes grados en la institución con el fin de fidelizar el personal y brindar un mejor servicio a la comunidad.

CONCLUSIONES

- Este trabajo permite concluir que la edad de retiro para la población de interés oscila entre los 6 y 10 años laborados, así mismo, las unidades de mayor deserción son las que presentan un alto grado de complejidad delictiva y riesgo para ejercer la labor policial.
- Los modelos de machine learning supervisados son apropiados para brindan confiablidad en la predicción de los datos y los resultados obtenidos confirman la factibilidad de utilización de las mismas.
- El mejor modelo para predecir los retiros de patrulleros en la Policía Nacional de Colombia es el XGBoost Extreme Gradient Boosting, con un 84.96% de exactitud, esto permite entrever el nivel de relevancia al momento de calificar en posible retiro de un funcionario que se encuentra laborando en la Policía Nacional en el grado de patrullero, teniendo como referente las variables propuestas para este estudio.

REFERENCIAS BIBLIOGRÁFICAS

Deo, R. (2015). Machine Learning in Medicine. Circulation, 132 (20), 1920-1930. https://www.ahajournals.org/doi/full/10.1161/CIRCULATIONAHA.115.001593

Abella, B. (2021). Mejora de las predicciones en muestras.

Sánchez, M. y García, M. (2017). Satisfacción Laboral en los Entornos de Trabajo. Una exploración cualitativa para su estudio. Scientia Et Technica, 22 (2), 161-166 https://revistas.utp.edu.co/index.php/revistaciencia/article/view/11471

B12. (2020, febrero 03). "*Qué es un modelo predictivo y cómo se aplica al negocio"*. https://agenciab12.com/noticia/que-es-modelo-predictivo-como-aplica-negocio#:~:text=Un%20modelo%20predictivo%20es%20un,vez%2C%20detectar%20oportunidades%20de%20negocio.

Dibble, S. (2001). *Conserve a sus empleados valiosos*. Oxford University.

Chitarroni, H. (diciembre de 2002). La regresión logística. *Instituto de Investigación en Ciencias Sociales IDICSO*. 1-8 https://racimo.usal.edu.ar/83/1/Chitarroni17.pdf

Naqa, I., Murphy, M (2015). ¿*Qué es el aprendizaje automático*? Springer, Cham.

Maguire, D. et al. (2005), *Spatial Analysis, and Modeling*. ESRI Press.

Giudici, P. (2003). *Applied Data Mining Stadistical Methods for Bussines and Industry*. Wiley.

Bhat , N., y McAvoy, T.J. (1995). Use of neural net s for dynamic modeling and cont rol of chemical process syst ems. Comput er Chem Engng, V.14(4-5), pp.573-583.

Adeniyi, DA, Wei, Z. y Yongquan, Y. (2016). Sistema de recomendación y minería de datos de uso web automatizado utilizando el método de clasificación K-Nearest Neighbor (KNN). Computación Aplicada e Informática, 12, 90-108. doi: 10.1016/j.aci.2014.10.001

Bolaños, C., Franco, f., & Marin, C. (2009). Diagnostico al proceso de retiro de personal en la Policía Nacional para planear estrategias que reduzcan los reintergros en la institución. Bogotá.

Bosco, J. (2020). Tutorial: XGBoost en Python.

Ceballos, n., & Campos, I. (2020). Perfil del retiro laboral temprano en población uniformada. https://repository.urosario.edu.co/bitstream/handle/10336/20957/PERFIL%20DEL%20RETIRO%20LABORAL%20TEMPRANO%20%20EN%20POLICIAS%20UNIFORMADOS%20DE%20LA%20POLICIA.pdf?sequence=1&isAllowed=y

Dempsey, J. R. (marzo de 1976). https://files.eric.ed.gov/fulltext/ED129920.pdf

Garza, F. (diciembre de 2019). Civinedu. http://www.civinedu.org/wpcontent/uploads/2019/12/CIVINEDU2019.pdf#page=267

Linares, H., Ramírez, A., & Camelo, G. (1984). Estudio exploratorio sobre el retiro a solicitud propia de agentes de la Policía Nacional por desmotivación. Bogotá.

Pimienta, J., Carvajal, J., & Calderon, j. (marzo de 2016). https://revistas.iue.edu.co/index.php/Psicoespacios/article/view/715/1013

Pinto, d., Malagon, D., Castro, J., Elizalde, M., & Redondo, P. (Diciembre de 2020). Politécnico Grancolombiano. https://alejandria.poligran.edu.co/bitstream/handle/10823/2616/JENNY%20TATIANA%20CASTRO.pdf?sequence=3&isAllowed=y

Reyes, G. (Noviembre de 2019). Dialnet. file:///C:/Users/LOPEZ/Downloads/Dialnet-PrediccionDeDesercionLaboralUtilizandoUnAlgoritmoG-7425952.pdf

Rodrigo, J. A. (2020). Máquinas de Vector Soporte (SVM) con Python.

Temple Univ Philadelphia Pa Merit Center. (Octubre de 1980). Defense technical Information Center. https://apps.dtic.mil/sti/citations/ADA093058

Terreros, A., Vega, A., & Pupo, J. (septiembre de 2019). Universidad Técnica de Machala. file:///C:/Users/LOPEZ/Downloads/Dialnet-AplicacionDelTeoremaDeBayesEnLaSeleccionDePersonal-7144063.pdf

Fidalgo F., y Gouveia, L. B. (2012). Turnover impact into organizational knowledge management. The case of employee turnover in Portuguese Real Estate. 6th Iberian Conference on Information Systems and Technologies (CISTI 2011), pp. 1-6.