

**CONTRIBUCIONES DESDE UN MODELO PREDICTIVO PARA IDENTIFICAR
EL PERFIL DE RIESGO DEL DEFRAUDADOR INTERNO EN UNA ENTIDAD
FINANCIERA DE COLOMBIA**

Autor:

Fabio Hernán Ruiz Galeano

Asesor:

Idier Albeiro Torres Guerra

UNIVERSIDAD EAFIT

Escuela de Administración - Departamento de Contaduría Pública

Maestría en Administración de Riesgos

2018

TABLA DE CONTENIDO

Resumen	1
Abstract.....	2
Introducción.....	3
1. Justificación.....	9
2. Objetivos del proyecto.....	14
2.1 Objetivo general	14
2.2 Objetivos específicos	14
3. Marco Teórico.....	15
3.1 El fraude.....	15
3.2 El fraude en Colombia.....	18
3.3 Técnicas para la prevención y detección del fraude.....	19
3.4 Perfil del defraudador	21
3.4.1 Empresas de auditoría	21
3.4.2 Otros investigadores.....	23
3.4.3 Comportamientos contraproducentes de los trabajadores	24
3.5 Técnicas de minería de datos	27
3.5.1 Modelos predictivos	28
4. Aspectos metodológicos.....	33
5. Resultados.....	37

5.1 Perfil global del defraudador a partir de estudios realizados por firmas de auditoría reconocidas mundialmente	37
5.2 Características de empleados asociados a eventos de fraude interno en la entidad financiera.....	41
5.2.1 Diagrama de la frecuencia de fraude por cargo	41
5.2.2 Histograma sobre la edad de los funcionarios defraudadores.....	44
5.2.3 Años de antigüedad del funcionario en la entidad financiera al momento de cometer el fraude o mala práctica.....	45
5.3 Evaluación comparativa de dos modelos estadísticos predictivos sobre el perfil de riesgo del defraudador interno en la entidad	46
5.3.1 Metodología para la evaluación del modelo predictivo	46
5.3.1.1 Modelos predictivos	47
5.3.1.2 Procedimiento para la validación	47
5.3.1.3 Procedimiento para la fragmentación de los datos.....	48
5.3.1.4 Validación Cruzada en Minería de datos	49
5.3.2 Modelo de Regresión Logística.....	50
5.3.2.1 Validación.....	52
5.3.2.2 AUC	53
5.3.3 Algoritmo Random Forest.....	54
5.3.3.1 Validación.....	55
5.3.3.2 AUC	56

5.3.4 Comparación de modelos y análisis.....	57
5.4 Aplicación del modelo Random Forest sobre la población de la entidad financiera para contribuir a crear su perfil de riesgo del defraudador interno.....	61
5.4.1 Características del perfil	61
5.4.2 Obtención del perfil de riesgo del defraudador mediante un árbol de decisión	63
6. Conclusiones	66
7. Referencias	71

ÍNDICE DE TABLAS

Tabla 1. Medición de las variables dicotómicas y categóricas	35
Tabla 2. Clasificación por cargo	43
Tabla 3. Matriz de confusión	48
Tabla 4. Resultados de la validación cruzada del modelo de regresión logística	58
Tabla 5. Matriz de confusión	58
Tabla 6. Matriz de confusión	55
Tabla 7. Comparación de modelos	57

ÍNDICE DE GRÁFICOS

Gráfico 1. Cómo se utiliza la tecnología para cometer fraude	16
Gráfico 2. Encuesta América Latina Vs Colombia 2016	39
Gráfico 3. Clasificación por cargo	42
Gráfico 4. Cargos de los trabajadores registrados en la base de infidelidad.....	49
Gráfico 5. Histograma de la edad de los funcionarios defraudadores	44
Gráfico 6. Histograma de la antigüedad de los funcionarios defraudadores	45
Gráfico 7. Validación	49}
Gráfico 8. Validación cruzada usando k grupos.....	50
Gráfico 9. Regresión logística – ROC Chart – AUC 0.64	59
Gráfico 10. All data	54
Gráfico 11. Variables importantes enmascaradas.....	55
Gráfico 12. Random Forest – ROC Chart – AUC 0.67	56
Gráfico 13. Simulación de la generalidad de los actos fraudulentos.....	60
Gráfico 14. Cargos.....	67
Gráfico 15. Género.....	67
Gráfico 16. Región.....	68
Gráfico 17. Jornada	68
Gráfico 18. Antigüedad	68
Gráfico 19. Dispersión de salarios	69

Resumen

Las empresas de servicios financieros son parte fundamental en la economía del país, y su continuo crecimiento ha ocasionado que sean más susceptibles al fraude debido a la falta de prevención y de controles que mitiguen las amenazas que van de la mano con la innovación tecnológica, los nuevos productos y las características del consumo. Esta problemática genera impactos negativos a nivel económico y social tanto para las empresas como para el país. Es por esto que se identificó la necesidad de contribuir a la elaboración del perfil de riesgo del defraudador interno en una entidad financiera que opera en Colombia, para ayudar en la prevención y detección de fraudes. Así, se definió un perfil parcial (puesto que solo contempla algunas variables) a partir del uso de un modelo estadístico predictivo (con funciones más probabilísticas que predictivas en el caso del fraude) que analizó la información proveniente de la caracterización de los empleados implicados en acciones fraudulentas, además de una revisión de literatura. A partir de la construcción del modelo de Random Forest, y con el objetivo de encontrar empleados con altas probabilidades de incurrir en actos incorrectos, se identificó como perfil defraudador de la entidad, los trabajadores que laboran en la red de sucursales, que poseen transacciones atípicas y en donde las relaciones de salarios vs los descuentos realizados al empleado no guardan una proporción. Los empleados que poseen cargos operativos, los cuales representan el 44.5 % de la empresa, son los más propensos a incurrir en fraude interno, el 66.4 % de los posibles defraudadores son mujeres; y las regiones del país que presentan mayor alerta son la región centro, con el 43.6 %, seguida de Bogotá y Sabana con el 28.7%

Palabras clave: Empleado, fraude interno, perfil de riesgo del defraudador interno, entidad financiera,

Abstract

Financial services companies are a fundamental part of the country's economy, and their continued growth has made them more susceptible to fraud due to the lack of prevention and controls that mitigate the threats that go hand in hand with technological innovation, new products and consumption characteristics. This problem generates negative economic and social impacts for both the company and the country. That is why it was identified the need to develop the risk profile of the internal fraudster in a financial entity, which allows the frauds prevention and detection from the characterization of employees involved in fraudulent actions, to achieve this, the mixed investigation was applied, which allowed the collection of data on fraud materialized by employees, their numerical measurement and respective statistical analysis to test the hypothesis, as well as a literature review. From the construction of the Random Forest model, and with the objective of finding employees with high probabilities of incurring in incorrect acts, the profile of the internal fraudster of the entity was identified, the employee who work in the branch network, who have atypical transactions and where the wage relationships Vs the discounts made to the employee do not keep a proportion. Employees who hold operational positions which represent the 44.5% of the company, are the most likely to incur internal fraud, 66.4% of possible fraudsters are women; moreover, country's regions that present greater alert are the central region with 43.6%, followed by Bogotá and Sabana with 28.7%

Key words: Employee, Internal fraud, risk profile of the internal fraudster, financial entity

Introducción

El fraude se refiere generalmente a algún acto intencional y doloso cometido por uno o varios individuos cuyo fin es la apropiación u obtención de una cosa ajena con el ánimo de dominio y uso. Según la página web de la ACFE (Association of Certified Fraud Examiners), el fraude es cualquier actividad o acción con el propósito de enriquecimiento personal a través del uso inapropiado de recursos o activos de una organización por parte de una persona.

De un modo más preciso, la ACFE usa el término “fraude ocupacional” en el Informe 2018 a las Naciones sobre Fraude y Abuso Ocupacional, donde analiza 2.690 casos de fraude ocupacional investigados entre enero de 2016 y octubre de 2017, con información proporcionada por los Examinadores de Fraude Certificados (ECF) que investigaron esos casos en 125 países. El fraude ocupacional es el uso de la ocupación de uno para enriquecimiento personal a través del uso indebido deliberado o mala aplicación de los recursos o bienes de la organización empleadora (ACFE, 2018).

Según ACFE (2018), las tres principales categorías ocupacionales de fraude son la malversación de activos, la corrupción y los informes fraudulentos de los estados financieros:

1. Apropiación de activos: puede ser el desvío de fondos en búsqueda de un beneficio propio.
2. Manipulación de la información financiera o contable: también llamada contabilidad creativa, como pudiera ser la manipulación de los resultados de la compañía para la consecución de determinados objetivos y obtención así del correspondiente bonus anual.
3. Corrupción: como sería la aceptación de tratos de favor en detrimento de otros para la obtención de un producto o servicio. Estaríamos hablando de conflictos de interés, sobornos, pagos inapropiados o, incluso, extorsión.

Las Normas Internacionales para el Ejercicio Profesional de la Auditoría Interna definen el fraude como cualquier “acto ilegal caracterizado por engaño, ocultación o violación de confianza, los cuales son perpetrados por individuos y por organizaciones para obtener dinero, bienes o servicios, para evitar pagos o pérdidas de servicios, o para asegurarse ventajas personales o de negocio”. (Institute of Internal Auditors [IIA], 2013, p. 10)

El fraude corporativo, despojado de su forma externa, no es más que el engaño, la mentira o el robo a una compañía. Más allá de este simple núcleo, el fraude puede clasificarse en dos tipos: fraude interno y fraude externo. En el presente estudio nos vamos a referir al fraude interno, que incluye todas aquellas formas en que un empleado puede engañar a una empresa para obtener un beneficio indebido. El fraude interno tradicionalmente implica apropiación indebida de activos y corrupción (Biegelman & Bartow, 2012), fenómenos a los que cualquier empresa puede estar expuesta. Por esta razón, si los controles internos en las organizaciones no están bien formulados y evaluados, o demuestran debilidad en sus procesos, existirá una alta probabilidad de que se materialice un fraude por parte de los colaboradores.

Lo anterior adquiere mayor importancia si se tiene en cuenta que las empresas pierden alrededor del 5% de sus ingresos por año debido a este delito (McMahon, Pence, Bressler & Bressler, 2016); el 75% de los fraudes los llevan a cabo los propios empleados de la compañía (PwC, 2014); y en relación con los efectos del fraude, son diversos los que se pueden presentar: costos directos e indirectos, afectaciones en la reputación de la empresa, debilidad en las relaciones con los grupos de interés, multas, sanciones y demandas; y a nivel macroeconómico, el fraude puede generar o incrementar fenómenos como el desinterés en los inversionistas, la desaceleración económica, la falta de empleo y la inflación.

Es relevante mencionar estos efectos del fraude a nivel macroeconómico por varias razones. Porque cualquier tema de investigación debe encuadrarse en un marco contextual más amplio que dé cuenta de su importancia y utilidad. Porque si bien no se profundizan las consecuencias macroeconómicas en la investigación, ellas evidencian la importancia del fenómeno del fraude, por el grado de incidencia mundial, que va más allá de cada empresa en particular. Y porque, aunque el perfil del defraudador no pueda ser el mismo para todas las empresas, el problema del fraude interno sí es generalizado en todas ellas, y en este sentido el tema se ubica en un contexto mayor.

Ninguna organización es inmune al fraude y a sus impactos negativos. En el caso de las entidades financieras, en el 2012 la presencia de fraude financiero fue del 7,60%, y en el año 2016, fue del 9,60% (ACFE, 2017). Por esto es necesario que las empresas cuenten con un programa de prevención de riesgos de fraude, que ayude a reducirlo con medidas de prevención, detección y respuesta; y que este programa sea visto no como un costo sino como una inversión, a pesar de que cuantificar el retorno de la inversión sea difícil (KPMG, 2016).

Como parte de estos programas y estrategias, los especialistas en fraudes llevan tiempo trabajando en la identificación de un perfil del defraudador interno que sea lo suficientemente preciso como para permitir a las organizaciones detectar a los autores durante la comisión del delito o incluso con anterioridad a que se produzca el fraude (KPMG, 2016).

A la hora de identificar y mitigar los riesgos, las empresas cuentan con la posibilidad de implantar herramientas eficaces de detección del fraude; pero muchas no han aprovechado todavía las posibilidades que las fuentes de información y las técnicas de Data analytics proporcionan para identificar y mitigar acciones fraudulentas (EY, 2016). Entre tales herramientas se encuentra, por ejemplo, la Minería de Datos (Data mining), que es un proceso iterativo de búsqueda de

información en grandes volúmenes de datos, para el modelaje de la información (López y Sánchez, 2012). Esta técnica también se define como el proceso de extraer conocimiento de cantidades masivas de datos (Pujari, 2012).

Las herramientas estadísticas basadas en redes neuronales artificiales y árboles de decisión para la detección de patrones de comportamiento han sido aplicadas con buenos resultados en campos muy heterogéneos (Muñiz y Álvarez, 1997; Olmedo, Velasco y Valderas, 2007), por lo que en este estudio se exploran las posibilidades que ofrecen estas herramientas estadísticas para el reconocimiento de patrones de fraude en operaciones individuales, ya que con las técnicas tradicionales se hace muy complicado cumplir estos objetivos (Dutta, 2013).

Es por eso esta investigación se busca dar respuesta a la pregunta ¿Cuál es el perfil parcial de riesgo del defraudador interno en una entidad financiera definido a través de un modelo estadístico predictivo, para contribuir a la prevención y detección de fraudes a partir de la caracterización de los empleados implicados en acciones fraudulentas?

En este sentido, el objetivo general de esta investigación es contribuir a elaborar el perfil de riesgo del defraudador interno en una entidad financiera de Colombia, aplicando un modelo predictivo para la prevención y detección de eventos de infidelidad. Es necesario aclarar que si bien se usa un modelo predictivo para lograr el objetivo de la investigación, en la comisión del fraude intervienen variables de tipo psicológico que impiden hablar de predicción en este fenómeno; por eso, el modelo, aunque en términos generales busca predecir, en el caso del fraude funciona para determinar probabilidades.

Ahora bien, para desarrollar el objetivo general, se realizará, inicialmente, una revisión de los perfiles globales del defraudador, a partir de los estudios aplicados por KPMG, PwC y Kroll; este ejercicio será de utilidad porque permitirá contar con diferentes aportes científicos confiables sobre

el defraudador, para comprender mejor su perfil y motivaciones; las causas y consecuencias de su actuación; las variables que se deben tener en cuenta para analizar el fenómeno, al igual que los métodos y las técnicas más viables. Posteriormente, se realizará una identificación de las características de los empleados relacionados con eventos de fraude materializados y/o malas prácticas en la entidad financiera; seguido de una prueba practicada, con los datos históricos sobre fraude en la entidad, a dos modelos estadísticos predictivos para definir cuál es el más efectivo en determinar el perfil de riesgo del defraudador en dicha entidad. Finalmente, se aplica el modelo elegido, corriéndolo sobre toda la población de la entidad, contribuyendo así a identificar el perfil de riesgo parcial del defraudador interno en la entidad financiera objeto de estudio.

Es de aclarar que determinada información sobre la entidad financiera no es revelada, pues está sujeta a un compromiso de confidencialidad. Así, ni el nombre ni otros datos que implícitamente podrían aclarar su identidad han sido consignados. Si bien datos como el número de oficinas de atención, el número de transacciones, los activos y otros más contribuirían al entendimiento de la operación, no es posible incluirlos porque son puestos por la entidad a disposición del público y, por lo tanto, si se consignaran, tácitamente permitirían determinar de qué entidad se trata.

También hay que precisar que los datos sobre el perfil psicológico de los empleados de la entidad son confidenciales y no se pudo tener acceso a ellos. Por esta razón no se incluyeron este tipo de datos en la creación del modelo predictivo. Además, el manejo de variables de esta clase requiere otro tipo de metodologías y una profundización en el área del comportamiento humano y la psicología que quedan más allá de los límites de esta investigación. Por todo esto, el perfil de riesgo del defraudador que se propone en esta investigación es parcial, sujeto al tipo de variables que se incluyen, y sabiendo que las variables psicológicas son muy importantes a la hora de

identificar un perfil más completo. Así, la investigación quiere contribuir a identificar ese perfil, pero no se propone ofrecer un perfil completo.

1. Justificación

Las empresas son parte fundamental de la economía de un país, de su estabilidad y desarrollo social, en especial las entidades financieras, las cuales otorgan préstamos, descuentos y anticipos, que promueven la creación, reorganización y crecimiento de las empresas (ANDI, 2017). Esta gestión se ve expuesta a diferentes problemáticas, entre ellas, el fraude, que evidencia el vacío que existe en cuanto a la regulación, control y monitoreo, debido a que la prevención del riesgo ha sido vista como un costo y no como una inversión, en razón de la dificultad para cuantificar el retorno de la inversión o para compararla con el costo del fraude. Bajo esta lógica, las compañías entienden que el fraude es un problema que puede conducir a pérdidas financieras y daños en la reputación, sin embargo, las empresas que tienen planes de cumplimiento y ética suelen incluirlos porque se sienten obligadas por alguna regulación (KPMG, 2016).

El fenómeno del fraude se adapta fácilmente a las características de cada organización, debido a las diferentes técnicas o métodos que puede utilizar, y porque el defraudador, al conocer la empresa, dispone de mayores oportunidades para explotar sus vulnerabilidades. Por su parte, la ACFE (2018), en su reporte a las Naciones Unidas sobre fraude y abuso ocupacional, indica que la pérdida conocida causada por el fraude a las organizaciones en todo el mundo durante el periodo de estudio fue de USD 7.1 mil millones; aunque la cifra es alta, no se acerca a la representación de la cantidad total de pérdida por fraude. La pérdida promedio debido a los fraudes fue de USD 2.75 millones, y la pérdida mediana fue de USD 130.000, que también es una cantidad enorme cuando se considera cuánto daño representa tal pérdida para la mayoría de las organizaciones.

No obstante, los efectos económicos no son los únicos que se pueden presentar, el fraude también genera afectaciones en la reputación a nivel local, nacional e internacional, influyendo negativamente en las relaciones comerciales con los diferentes grupos de interés de la compañía,

especialmente los clientes, proveedores e inversionistas, quienes tendrán una imagen de la empresa relacionada con la falta de seguridad (Asociación Española de Empresas contra el Fraude [AEECF], 2017).

Al respecto, los resultados de la Encuesta sobre el Perfil del Defraudador de KPMG 2016, indican que en Colombia el 34% de las empresas encuestadas manifestó haber tenido como consecuencia del fraude un incremento en los costos de operación, el 23% señaló haber experimentado distorsión en los precios del mercado, el 19% vivenció daños en las relaciones comerciales, el 14% indicó que se presentaron afectaciones en la imagen corporativa y el 10% manifestó que les ocasionó multas, sanciones o demandas.

Algunos de los anteriores efectos son mencionados también por Kroll a través de su encuesta global de fraude 2016, en la que señala que como consecuencia del fraude se presenta mala reputación de la empresa, a nivel interno y externo, y el grado de impacto depende del manejo que se le dé a la problemática; así mismo, señala una disminución en los ingresos de las entidades afectadas por fraude de entre el 1% y 3%, y pérdida e insatisfacción de sus clientes; además, indica cómo se ve reflejado el fraude en la afectación a la moral de los trabajadores, y que este fenómeno debe ser objeto de mayor control e investigaciones detalladas por diferentes instituciones.

Por su parte, la ACFE, en su Estudio Global sobre Fraude 2016, indica que la pérdida total causada por el fraude excedió 6.3 billones de dólares, con una pérdida promedio por caso de 2.7 millones de dólares. La mediana de pérdida para todos los casos del estudio fue 150,000 dólares, con 23.2% de los casos causando pérdidas de 1 millón de dólares o más.

Es decir, entonces, que la realidad económica del sector financiero y los riesgos a los que está expuesto han motivado la investigación desde un enfoque centrado en los impactos y consecuencias que tiene el riesgo de fraude en las empresas, con el objetivo de sensibilizar a los

empresarios sobre la necesidad de tomar decisiones respecto a la prevención, control y monitoreo del fraude interno.

A pesar del contexto informativo que existe sobre esta problemática, aún se evidencia un vacío de conocimiento en el mundo y en Colombia sobre estudios de modelos o perfiles del defraudador interno, y la información de las encuestas realizadas por empresas de auditoría como KPMG, PwC y Centro de estudios EY y Kroll arrojan datos que son confiables en el universo de cada investigación pero que no convergen, pues hay discrepancias entre sí en cuanto a los cargos, edad, antigüedad, salario, género, jornada, entre otras variables de los perfiles del defraudador interno. Además de estas discrepancias, los resultados de las encuestas no son aplicables a la realidad en Colombia y a la de la empresa objeto de estudio, debido a las diferencias en las características sociales, culturales, económicas, tecnológicas y demográficas que están presentes en un país y en una empresa. Y en este último sector también influyen la razón social, el tamaño, los cargos y ocupaciones, las características demográficas, sociográficas, psicográficas, laborales, conductuales y transaccionales de los trabajadores. Es por esta situación, precisamente, que los resultados de las empresas de auditoría y de los pocos autores que han investigado el tema (Arcenegui, Obrero y Martín, 2015; Ayala, 2008; Calleja, 2014; Ramírez, 2008), a pesar de que utilizan variables similares para evaluar, obtienen resultados diferentes, que pueden ser confiables pero no garantizan que con base en esa información se pueda interpretar la realidad del fraude interno en un país diferente al de cada investigación o en una determinada empresa. Es decir, el perfil del defraudador no está establecido.

Por consiguiente, no es fácil clasificar el fraude financiero y determinar sus causas precisas, pues usualmente el fraude no es realizado por una sola persona, ni en un mismo instante, sino que surge por varios factores que se traslapan en la perpetración del mismo.

Este proyecto permite una aproximación a la aplicación de un modelo predictivo para la prevención y detección de eventos de infidelidad, utilizando variables que “mediante la combinación de sus características puedan ser clasificadas como fraude, ayudando a priorizar los recursos de investigación disponibles hacia las empresas que concentran mayor potencial de fraude” (Badal y García, 2016, p. 3).

Esta situación será objeto de análisis en una entidad financiera de Colombia que aporta a la economía, el desarrollo y la estabilidad del país, y cuyo crecimiento hace que sea más vulnerable frente a la presencia del fraude; además de que esta problemática le genera a la entidad impactos negativos en los ingresos, los costos directos e indirectos, el interés por parte de los inversionistas, la reputación y las relaciones con los grupos de interés.

El efecto negativo del fraude en la imagen de la empresa puede realmente afectarla: una entidad financiera que es objeto de fraude pierde uno de sus pilares, a saber, la imagen de confianza y estabilidad. Inclusive, en caso de que el fraude afectara a los clientes, aunque el banco les reconociera las pérdidas debidas al fraude, al percibir que la entidad no es segura, habría una tendencia a retirarse o prescindir de sus servicios.

Así mismo, esta investigación ofrecerá elementos para que la empresa pueda mejorar en la gestión preventiva o la detección del fraude de forma oportuna. Actualmente, esta entidad solo detecta los fraudes una vez materializados, es decir, se han evitado pérdidas mayores al detener fraudes en proceso, pero no se han detectado antes de investigarlos como producto de alguna alerta. Con los datos arrojados por la investigación puede iniciarse realmente la detección del riesgo.

Además, al ayudar a identificar el perfil de riesgo del defraudador, se puede contribuir a priorizar los recursos de investigación disponibles hacia las personas que concentran mayor potencial de fraude. Ya que la entidad no realiza ningún proceso para la detección del riesgo de

fraude interno, contar con una fuente de información que pueda alertar sobre la probabilidad de fraude, ayuda a determinar hacia dónde se deben enfocar los esfuerzos de investigación, de este modo, se van definiendo puntos críticos de análisis y las necesidades de recursos humanos, técnicos y financieros para evitar que ocurran los fraudes.

Estos resultados también contribuirán al crecimiento de la compañía financiera. Incluso, los resultados de esta investigación pueden ser analizados y puestos en práctica por otras entidades bancarias de características similares en el país.

2. Objetivos del proyecto

2.1 Objetivo general

- Contribuir a la elaboración del perfil de riesgo del defraudador interno en una entidad financiera de Colombia, a partir de la caracterización del riesgo expresado por empleados y la aplicación de un modelo predictivo para la prevención y detección de posibles eventos de infidelidad.

2.2 Objetivos específicos

- Revisar el perfil global del defraudador, a partir de estudios realizados por firmas de auditoría reconocidas mundialmente, como KPMG, PwC y Kroll.
- Identificar las características de empleados asociados a eventos de fraude interno en la entidad financiera.
- Elegir mediante pruebas un modelo estadístico predictivo para aplicarlo sobre la población total de la entidad financiera e identificar así parcialmente su perfil de riesgo del defraudador interno.

3. Marco Teórico

3.1 El fraude

La ACFE comprende el fraude como aquellas actividades o acciones que se realizan con el objetivo de obtener un enriquecimiento personal por medio del uso inapropiado o la sustracción de recursos o activos de una empresa (ACFE, 2017).

En cuanto a los tipos de fraude, ACFE (2017) afirma que puede tener implicaciones internas y externas, “internamente por empleados, directivos, funcionarios o propietarios de la empresa, o externamente por los clientes, proveedores y otras partes” (párr. 3).

En esta investigación el enfoque estará en el fraude interno, que también es conocido como fraude ocupacional o laboral, y se refiere a aquél en donde los trabajadores se aprovechan de su empleo para obtener un enriquecimiento personal, haciendo un mal uso deliberado o una indebida aplicación de los recursos o activos de la empresa. Es decir, el empleado comete fraude contra su empleador. Este tipo de acto se caracteriza en la actualidad por apoyarse en la tecnología y en metodologías que permiten ocultar las acciones fraudulentas; no obstante, al final la falta de ética del defraudador puede ser probada a lo largo del tiempo (ACFE, 2017).

El estudio de KPMG (2016) “sugiere que la tecnología se utiliza con mayor frecuencia en perpetrar el fraude que en detectarlo. La tecnología era un habilitador importante para el 24 % de los defraudadores” (p. 20), quienes afirmaron que se apoyaron en la tecnología para crear información falsa o engañosa en los registros contables, seguido del 20%, que proporcionó información falsa o engañosa a través de correo electrónico u otra plataforma de mensajería, mientras que el 13% abusó del acceso a los sistemas informáticos (ver Gráfico 1).

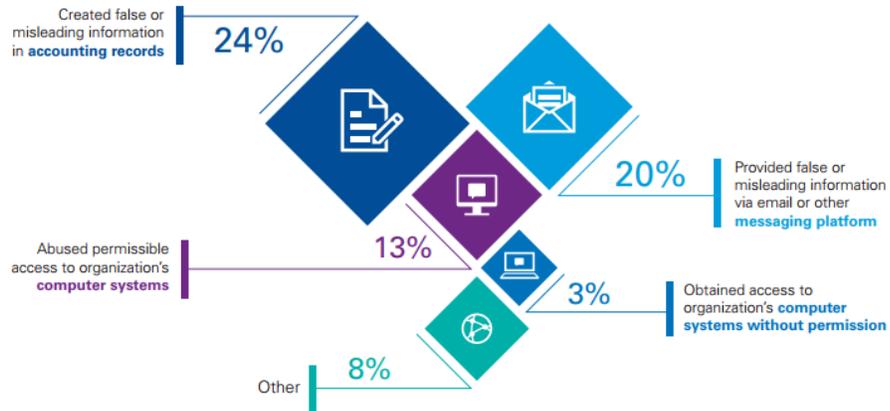


Gráfico 1. Cómo se utiliza la tecnología para cometer fraude

Fuente: KPMG. (2016). Global Profiles of the fraudster.

Adicional a esto se conoció que el 60% de los fraudes relacionados con tecnología implican a los jóvenes entre los 26 y 45 años, mientras que los defraudadores mayores se inclinan más por las relaciones personales (KPMG, 2016).

Para explicar por qué ocurre el fraude puede aplicarse el modelo del triángulo, propuesto por el criminólogo Donald Cressey en una investigación enfocada en las personas que se apropian de dinero o bienes públicos, a quienes llamó “violadores de confianza”. Cressey señala que existen tres condiciones que deben estar presentes en un fraude: la oportunidad para cometer fraude debido a los controles ausentes o ineficaces, un incentivo o presión económica, y una racionalización de la conducta fraudulenta (Kassem y Higson, 2012).

Esta información es asumida como parte de las características del fraude por el Instituto de Contabilidad y Auditoría de Cuentas (ICAC), en su Norma Internacional de Auditoría 240, sobre las responsabilidades del auditor en la auditoría de estados financieros. Con respecto a la oportunidad, la ICAC señala que se presenta cuando el trabajador considera que el control interno puede evadirse, “por ejemplo, por encontrarse esa persona en un cargo de confianza o porque

conoce deficiencias específicas en el control interno” (2013, p. 12); respecto a la presión, un ejemplo es el de la dirección, cuando desde una instancia interna o externa a la entidad se ve impulsada a lograr un objetivo de beneficios o resultados financiero, y especialmente si se perdiera una ventaja significativa para la dirección dado el caso de no conseguir tal objetivo. Finalmente, en cuanto a la racionalización, esta se relaciona con aquellos trabajadores cuyos valores éticos no les impiden cometer actos deshonestos ni tener una actitud crítica frente a ellos, actuando así consciente e intencionadamente; “sin embargo, incluso personas que en otra situación serían honestas pueden cometer fraude en un entorno que ejerza suficiente presión sobre ellas” (ICAC, 2013, p. 12).

El modelo del triángulo de Cressey es una buena práctica en auditoría, destinada a descubrir fraudes. Este modelo ha inspirado algunas variantes, como el modelo del diamante, en el cual, además de las tres condiciones enunciadas, se une la competencia del defraudador (Wolfe y Hermanson, 2004), referida a la oportunidad relacionada con el conocimiento o posición del defraudador.

Al respecto, Olcina (2016) afirma que el llamado Triángulo del Fraude también pone de manifiesto que los tres factores no suelen manifestarse de forma independiente y aislada; al contrario, detrás de la mayoría de fraudes ocurren dos o tres de los factores de forma combinada.

Finalmente, se conoce que la información financiera fraudulenta puede obtenerse mediante:

- La manipulación, la falsificación o la alteración de los registros contables o de la documentación de soporte a partir de los cuales se preparan los estados financieros.
- El falseamiento o la omisión intencionada de hechos, transacciones u otra información significativa en los estados financieros.

- La aplicación intencionadamente errónea de principios contables relativos a cantidades, a la clasificación, a la forma de presentación o la revelación de la información (ICAC, 2013, p. 12).

3.2 El fraude en Colombia

El fraude en Colombia es una problemática que está en aumento, y cada vez se descubren más casos de mala ética y falta de integridad, principalmente en el sector salud, financiero y bancario. Este fenómeno produce pérdidas millonarias que pueden llevar a la quiebra a una entidad, y a nivel general, obstaculizar el emprendimiento. Según KPMG, en su última encuesta de fraude realizada para Colombia (2013), aproximadamente 7 de cada 10 empresas en el país habían sido víctimas de al menos un fraude en el último año, y el 46% de estos casos estuvieron relacionados con la malversación de activos, mientras que el 51%, con el fraude financiero, generando un daño económico estimado de 3.600 millones de dólares para 2013.

En la actualidad, el país es muy susceptible frente al fraude. Para el 2013, según KPMG, de los 197 directivos de empresas encuestados, el 65% de las empresas en Colombia padeció fraude, pero todo indica que la cifra va en aumento, pues en la encuesta de Kroll 2015-2016, se encontró que el país tenía la incidencia general de fraude más alta de entre 8 países encuestados, entre ellos, México, Brasil, India y Estados Unidos. Esta incidencia fue del 83% frente al 63% de su encuesta 2013-2014. Kroll también halló que Colombia fue el único país que registró un crecimiento en la pérdida promedio por fraude: para la encuesta anterior fue del 0,7%, para la última subió al 0,9%. Llama también la atención que en esta encuesta Colombia es el país que más presenta robo o pérdida de información o ataque informático, con un 27%. Estos datos indican que el problema del fraude en el país es creciente y que empieza a ser relevante la ciberdelincuencia.

3.3 Técnicas para la prevención y detección del fraude

El fraude, al ser una problemática con trascendentales impactos para las empresas, la economía y el bienestar de la sociedad, ha sido analizado por diferentes instituciones. Una de ellas es el ICAC, que por medio de la Norma 240 comparte las principales técnicas aplicadas por la dirección para cometer fraude, evadiendo controles a través de la elaboración de registros de asientos ficticios en el libro diario, ajustando hipótesis y cambiando los juicios para el cálculo de saldos contables. Otras técnicas usadas son:

- Omitir, anticipar o diferir el reconocimiento en los estados financieros de hechos y transacciones que han ocurrido durante el periodo.
- Ocultar, o no revelar, hechos que podrían afectar a las cantidades registradas en los estados financieros.
- Realizar transacciones complejas estructuradas de forma que falseen la situación financiera o los resultados de la entidad.
- Alterar las condiciones y los registros relativos a transacciones significativas e inusuales (ICAC, 2013)

Teniendo en cuenta las anteriores situaciones, las empresas deben tomar medidas al respecto que les permitan un mayor control y monitoreo. La auditora forense Muna Buchain (2013) sugiere técnicas como el fortalecimiento del control interno y la evaluación de riesgos, el análisis y las revisiones forenses en áreas críticas, la segregación de funciones y tramos de responsabilidad, el diseño y promoción de programas antifraude, el aumento en la seguridad de los sistemas de información, el establecimiento de mecanismos para reportar fraudes, la verificación de los antecedentes laborales y la aplicación de pruebas de confianza a los trabajadores.

Por su parte, KPMG (2016) sugiere que para combatir el fraude en las organizaciones es necesario analizar los riesgos de fraude de forma recurrente y proactiva, utilizar recursos tecnológicos como los sistemas de monitorización de amenazas y *Data & Analytics* (D&A), que contribuyen a la identificación de los comportamientos anómalos o sospechosos. Así mismo, señala que se debe conocer y seleccionar a los terceros con los que se relaciona la organización, atendiendo a criterios de ética e integridad, y permanecer alerta para afrontar las amenazas internas. Adicional a esto, sugiere que es necesario contar con un perfil del defraudador, pues con este es posible “diseñar una línea de defensa corporativa para detectar, responder y mitigar los posibles casos de fraude que pueden presentarse en una empresa e igualmente permite asignar eficientemente los recursos disponibles a las áreas o funciones de mayor riesgo” (KPMG, 2016, párr. 5).

Para Mondragón (2011), en la actualidad la tendencia en prevención y detección del fraude está muy relacionada con el uso de sistemas analíticos y predictivos, que influyen en el buen desempeño de herramientas calificadoras del fraude. Tal es el caso de la minería de datos, la cual:

Reúne técnicas provenientes de diversas disciplinas como la estadística, la inteligencia artificial, la computación gráfica y el procesamiento masivo de datos y que ayudan de manera rápida y fácil a desarrollar las reglas para definir un patrón de comportamiento del fraude para identificar y detener las transacciones de este tipo al momento que son hechas en el punto de venta (2011, párr. 12)

Por otro lado, también existen modelos de *Machine Learning* (ML), que ayudan a la prevención del riesgo. Uno de ellos es el de regresión logística, que contribuye a clasificar y facilitar la interpretación de los coeficientes, para poder determinar relaciones entre variables explicativas y la respuesta, es decir, hacer inferencia. También existen los árboles de decisión, los cuales son un

modelo no paramétrico que ayuda a establecer las reglas de asociación entre las variables del modelo, además de ser clave para las predicciones. Otro más es el de los bosques aleatorios, que “predicen muy bien, pero no permiten la interpretación del modelo, utilizan remuestreo (son computacionalmente exhaustivos, por eso son buenos con la predicción)” (R Notebook, 2016, p. 2).

3.4 Perfil del defraudador

El ciudadano que comete fraude en una empresa tiene diversas características a nivel psicológico, sociológico, educativo y de riesgo conductual, que pueden influir en las motivaciones y toma de decisiones, así como en la forma de actuar y de interpretar lo que sucede a su alrededor. Al respecto es necesario especificar que el enfoque del perfil parcial del defraudador interno de la entidad financiera objeto de estudio que se quiere establecer se realizará a partir de una caracterización de las variables ocupacionales y sociodemográficas de los defraudadores.

3.4.1 Empresas de auditoría

Según KPMG (2013), el perfil del defraudador es difuso, ya que el fraude es una problemática que puede ocurrir en cualquier organización, y cada empresa requerirá un análisis personalizado que considere sus características, los cargos y los empleados con los que cuenta.

En la encuesta de *Perfiles globales del defraudador* (2016), en donde se analizó a 750 defraudadores de 78 países, KPMG indicó que el mayor porcentaje de agresores internos corresponde a los trabajadores del nivel ejecutivo o directores no ejecutivos con el 34%, el 32 % son gerentes y el 20 % son miembros del personal. También encontró que el 79% de las personas que cometen fraude son hombres, mientras que las mujeres tienen un 17% de participación. En cuanto a la edad, se conoce que el 37% tiene entre 36 y 45 años, seguido del 31%, que tiene entre

46 y 55 años. Respecto a la antigüedad en la empresa, el 38% tiene 6 años o más en ella, el 19% tiene de 1 a 4 años y el 14% de 4 a 6 años.

Para el caso de Colombia, las encuestas más recientes son las de PwC (“Encuesta de Delitos Económicos 2016, Capítulo Colombia”), con 118 participantes de Colombia, entre directivos y jefes de departamentos; y Kroll (“Informe Global de Fraude 2015/16), que entrevistó a 768 altos ejecutivos de todo el mundo. Según PwC, en Colombia los cargos gerenciales principiantes son los más propensos a cometer fraude (61%), seguidos por los cargos intermedios (35%) y los directivos (4%), en oposición a las cifras globales de su encuesta, en donde los más propensos son los mandos medios; así mismo, la edad de los defraudadores está entre los 31 y 40 años (igual al promedio en Latinoamérica), y el tiempo de antigüedad en la empresa se encuentra entre los 6 y 10 años, mientras que en América Latina está entre los 3 y 5 años (2016). Por su parte, los resultados de la encuesta de Kroll (2015-2016), arrojan que casi la mitad de los fraudes en Colombia los cometieron ejecutivos sénior o la gerencia media, con el 44%, mientras que a nivel global son los empleados junior los que encabezan sus encuestas, con el 45%, frente a un 36% para gerencia media (esta encuesta no presenta datos sobre edad ni antigüedad en la empresa).

Como puede verse, las diferentes encuestas presentan ciertas divergencias, pero las más importantes se refieren al tipo de cargos que se implican más en el fraude. Estas divergencias pueden deberse al sesgo de los entrevistados. Mientras KPMG investiga a los propios defraudadores (y por lo mismo su encuesta es de mucha utilidad para estudiar sus perfiles), PwC entrevista a directivos y jefes de departamentos, y Kroll encuesta a altos ejecutivos. Cada encuesta genera un sesgo propio, incluida la de KPMG, pues muy bien puede haber seleccionado, entre los millones de casos de fraude comprobado, aquéllos más relevantes. Las numerosas investigaciones y esfuerzos que se han realizado para establecer el perfil del defraudador han sido importantes

porque generan conocimiento valioso para la toma de decisiones; no obstante, la discrepancia en algunos de los resultados es una evidencia de que el perfil debe ser evaluado y analizado según las características de cada país y organización, para generar mayor confiabilidad.

3.4.2 Otros investigadores

Los aportes sobre el perfil del defraudador que han realizado empresas de auditoría son insumos valiosos y confiables que permiten la comprensión de la problemática. Pero también diversos autores se han interesado por el tema y, a través de una metodología rigurosa, han realizado aportes al respecto, incluso, usando las citadas encuestas como base argumentativa.

Entre ellos, Calleja (2014), estudiando el caso español, indica que los defraudadores son generalmente *mutantes*, porque implementan medios novedosos que son difíciles de identificar; esto hace que el fraude no sea un concepto establecido, al igual que el perfil de quien lo ejecuta. Por otro lado, esta autora señala que los defraudadores en España son individuos con formación profesional y alto status social.

Por otro lado, Ayala (2008) señala que la edad promedio del defraudador está entre los 36 y 55 años, la antigüedad en la empresa es mayor a seis años, y desempeñan cargos en el área financiera, contable y de control de gestión, principalmente. En cuanto a la frecuencia según el sexo, manifiesta que, con un 61%, los hombres son quienes tienden a cometer un delito de este tipo, situación que se presenta precisamente porque son los que ocupan los altos cargos gerenciales, mientras que las mujeres no lo hacen. Así, ya que los defraudadores suelen ser personas claves en la administración de la información confidencial y financiera, se les facilita evadir, ocultar y engañar los controles internos.

Ramírez (2008) postula que el fraude ha evolucionado hasta el punto de crear al empleado “tecnofraucrata”: una persona preparada, con buen manejo de los negocios, de los mercados, los procedimientos, los sistemas; y que además tiene acceso a la última tecnología para falsificar información física y electrónica.

Finalmente, para Arcenegui, Obrero y Martín (2015), el perfil característico del defraudador en banca paralela se fundamenta en que trabaja para un banco, situación que facilita el acceso a los inversionistas para obtener su confianza y dinero; también, maneja efectivo, es mayor a los 50 años de edad en el 61,1% de los casos, y tiene una antigüedad de más de 10 años en la empresa en un 47,2%. En relación con el cargo que ocupa, el 80% asume la responsabilidad de director, y el 7% de subdirector y empleado de oficina. Adicional a esto, los empleados que representan un mayor riesgo de cometer fraude no muestran inconvenientes en ganar dinero sin importar los medios necesarios para obtenerlo.

3.4.3 Comportamientos contraproducentes de los trabajadores

El término “Comportamientos contraproducentes en el trabajo” (CWB) se refiere a los comportamientos voluntarios de los empleados, que tienen el potencial de dañar a una organización, a sus miembros o a ambos (Spector y Fox, 2005). En este sentido, Robinson y Bennett (1995) los definieron como comportamientos antisociales ejecutados por los empleados en su lugar de trabajo, siendo potencialmente dañinos para la organización. Entre ellos se incluyen la rudeza, la pérdida de tiempo, el esfuerzo para retener y el sabotaje (Penney, Hunter y Perry, 2011).

Hay investigadores que se han enfocado en el perfil del defraudador desde los comportamientos contraproducentes, que, según Omar, Vaamonde y Delgado (2012), son las acciones voluntarias y

discrecionales que vulneran las normas establecidas por la organización. Bolton, Harvey, Grawitch y Peluquero (2012) señalan que la sensación crónica de agotamiento físico y emocional, el bajo y moderado autocontrol, la despersonalización que corresponde al desprendimiento de la persona con el trabajo, la insensibilidad emocional en relación con los clientes y compañeros, la disminución de la realización personal, además de la oposición a una organización cognitiva, influyen en los comportamientos contraproducentes en el trabajo.

Este enfoque en el origen emocional del comportamiento contraproducente es respaldado por Penney *et al.* (2011), quienes estudian cómo la escrupulosidad y la estabilidad emocional (dos de los cinco grandes rasgos de personalidad) están asociadas con la gestión estratégica de recursos (los objetos, características personales, condiciones o energías que el individuo valora o mediante los cuales puede conseguir estos recursos valiosos) en el trabajo. Las autoras afirman que los empleados manejan mejor sus recursos y se abstienen de comportamientos contraproducentes cuando presentan no solo una conducta concienzuda, orientada hacia logros y disciplinada, sino que además son estables emocionalmente y, por lo mismo, no tienen que invertir su energía en controlar sus emociones negativas. Por añadidura, las organizaciones deben proveer los recursos adecuados para que los empleados desempeñen sus labores; de este modo, se reduce la presencia de los comportamientos contraproducentes.

En otras palabras, fenómenos como el agotamiento y la inestabilidad emocional (que hacen a las personas ansiosas, impacientes, emocionales, inseguras, con niveles altos de preocupación y facilidad para el enojo y la irritación) pueden generar depresión y desesperanza, e influir en que el trabajador no esté orientado hacia el cumplimiento de objetivos sino hacia evitar el fracaso, contrario a quienes son estables y, por ende, tienen menos necesidad de invertir tiempo en regular

sus emociones, lo cual genera que tengan una mayor capacidad de asignar recursos para realizar las tareas.

A lo anterior se le suma lo planteado por Fida, Paciello y Tramontano (2014), para quienes el comportamiento contraproducente está relacionado con el estrés causado por factores que están en la organización y que generan frustración, afectando la consecución de los objetivos laborales, la realización de actividades y el rendimiento en el trabajo. Estas situaciones generan sentimientos negativos y comportamientos inadecuados, como una estrategia para reducir la condición emocional negativa causada por las frustraciones experimentadas en la empresa. Este estrés está relacionado con la falta de libertad en la toma de decisiones y la carencia de apoyo social, variables que se relacionan con la carga de trabajo. Por otro lado, las emociones negativas, como la ansiedad e irritabilidad, son otros factores resultantes de situaciones percibidas como negativas y que generan niveles bajos de satisfacción laboral. Finalmente, la desvinculación moral también está presente como un comportamiento ilícito, desviado y antisocial, a través de esta situación:

El contenido moral y lazos por lo general asociados con el acto desviado se separan (o desacoplan) de él, y por lo tanto, llevar a cabo ese comportamiento en la búsqueda del deseo o la meta propia no se considera ni internamente aversivo ni obstructivo por el actor. En cierto sentido, es un proceso por el cual el individuo puede racionalizar, por cualquier excusa o justificación, el daño y el mal al que el comportamiento sirve necesariamente, de manera que se neutralizan los mecanismos de disuasión (por ejemplo, la culpa) con respecto al comportamiento están neutralizados (2014, p. 133).

Por su parte, Omar *et al.* (2012) asocian el estrés laboral con la sobrecarga de trabajo, el exceso de tareas, la ambigüedad del rol y el agotamiento; las percepciones de injusticia a nivel distributivo: en el salario, los premios, las promociones, etc.; procedimental: en las normas y procedimientos;

e interpersonal, en aspectos como la exclusión, ser ignorados, rechazados por los compañeros de trabajo y el supervisor, situaciones que además puede generar que el trabajador realice sabotaje, abuse del alcohol y disminuya la producción. La deseabilidad es otro de los comportamientos condicionados por los trabajadores que buscan mostrar una imagen mejorada de sí mismos; y, finalmente, los datos sociodemográficos relacionados con la edad, el sexo, el nivel de educación, el cargo y la trayectoria laboral.

Finalmente, es importante mencionar la postura de Grijalva y Newman (2015), para quienes el comportamiento contraproducente está muy relacionado con el narcisismo, presente en las personas que tienen altos niveles de preocupación sobre la importancia personal, requieren de admiración y presentan falta de empatía; además, que se vuelven agresivas cuando su autoestima se encuentra amenazada y experimentan cólera, generando una tendencia a los comportamientos que pueden afectar a la empresa. No obstante, “el narcisismo es menos fuertemente relacionado con el comportamiento laboral desviado en las culturas que tienen valores colectivistas (es decir, las culturas que tienen un especial interés en las relaciones cohesivas y la lealtad)” (p. 119).

3.5 Técnicas de minería de datos

Dentro de la minería de datos se distinguen tipos de tareas, cada una de las cuales puede considerarse como una clase de problema a ser resuelto por un algoritmo de minería de datos. Las distintas tareas pueden ser predictivas o descriptivas. Entre las tareas predictivas encontramos la clasificación y la regresión.

La clasificación es quizá la tarea más utilizada. En ella, cada instancia (o registro de la base de datos) pertenece a una clase, la cual se indica mediante el valor de un atributo que se llama la clase

de la instancia. Este atributo puede tomar diferentes valores discretos, cada uno de los cuales corresponde a una clase (Ramírez, Hernández y Ferri, 2007).

La regresión es también una tarea predictiva que consiste en aprender una función real que asigna a cada instancia un valor real. Esta es la principal diferencia con respecto a la clasificación; el valor a predecir es numérico. El objetivo en este caso es minimizar el error entre el valor predicho y el valor real (Ramírez *et al.*, 2007).

En la minería de datos existen diversas técnicas, dado que es un campo muy interdisciplinar. Entre las más utilizadas se encuentran los árboles de decisión, las redes neuronales, el aprendizaje bayesiano, la regresión logística y el Random Forest, entre otras.

3.5.1 Modelos predictivos

Los modelos predictivos permiten valorar la relación que existe entre conjuntos de variables para posibilitar la toma de decisiones confiables. A través de ellos se pueden estimar valores futuros de variables de interés, que predigan las probabilidades de que una persona, según ciertas características, reaccione de una forma determinada. “Una vez ingresados los datos de la persona y se aplique el modelo predictivo se obtendrá una calificación que indicará la probabilidad de que se produzca la situación estudiada por el modelo” (Espino, 2017, p. 9). Es decir:

Un modelo predictivo es un mecanismo que predice el comportamiento de un individuo. Utiliza las características del individuo como entrada y proporciona una calificación predictiva como salida. Cuanto más elevada es la calificación, más alta es la probabilidad de que el individuo exhiba el comportamiento predicho. (2017, p. 12)

El término *calificación* en este contexto hace alusión a las tendencias y posibilidades presentes en un grupo, “pero no garantiza que la predicción se cumpla en cada caso individual, pues una

probabilidad individual por naturaleza simplifica excesivamente la cosa del mundo real que describe” (Espino, 2017, p. 12). Así, se debe tener en cuenta que un modelo predictivo nunca será 100% confiable; incluso habrá ocasiones en las que se aleje de los resultados óptimos, debido a que por muy frecuentemente que se haya identificado un patrón de comportamiento, este no tiene necesariamente que repetirse. A pesar de la anterior afirmación, en el contexto empresarial, y teniendo en cuenta los impactos negativos del fraude como problemática, siempre será mejor intentar predecir y tomar decisiones con base en un modelo, y no, por el contrario, tratar de adivinar sin fundamentos (2017).

El análisis predictivo construye un modelo estadístico que requiere tener unos datos existentes, unos atributos o comportamientos conocidos para predecir datos que se espera conocer pero de los que no se dispone (2017).

Por su parte, en el blog Logicalis (2015), se afirma que:

Los modelos predictivos, para poder llevar a cabo su misión, requieren de predictores y de la observación de los conjuntos de datos. A mayor número de predictores y mayor profundidad en su investigación, aumentará la complejidad del análisis. Aunque éste no es el reto más complicado. El verdadero desafío para los modelos predictivos es encontrar buenos subconjuntos de predictores o variables explicativas, es decir, hallar los que mayor utilidad aportan, los que mejor se ajustan a los datos (párr. 8).

Así mismo, en los modelos predictivos se resaltan las oportunidades que generan para el análisis de una realidad basándose en la relación de unas varias variables con otras. Este ejercicio implica gestión, tecnologías de la información y modelado, gran cantidad de datos y tener como objetivo aportar al éxito en los negocios por medio de la toma de decisiones. Este modelo es

aplicado en contextos de inversión, previsión de ventas, cuando se busca conocer segmentos rentables y en la identificación de riesgos (Logicalis, 2015).

En el contexto de los negocios, en Logicalis (2015) se resaltan como beneficios de la aplicación de este modelo la optimización de la toma de decisiones, la minimización del riesgo, poder conocer mejor al cliente, predecir acciones futuras, mejorar la consistencia de los negocios y la satisfacción del cliente, reducir costos y potenciar la competitividad.

Para el modelado de la información se utilizan técnicas como Random Forest (Bosques Aleatorios) y Regresión Logística. En esta investigación, como técnica de los modelos predictivos, se aplicará el Random Forest o bosque aleatorio, que combina árboles de decisión independientes que tienen la capacidad de analizar diversos datos de muestra, de evaluar diferentes variables:” los resultados de los diversos árboles de decisión se convierten entonces en un promedio simple o agregados a través de una ponderación adicional” (Espino, 2017, p. 16). En cuanto a la construcción de esta técnica, es necesaria la creación de numerosos árboles de decisión que sean independientes gracias a datos de entrada distintos. Así, se altera el conjunto inicial de partida, del modo siguiente:

- Se selecciona aleatoriamente con reemplazamiento un porcentaje de datos de la muestra total.

Es habitual incluir un segundo nivel aleatoriedad, esta vez afectando los atributos:

- En cada nodo, al seleccionar la partición óptima, tenemos en cuenta sólo una porción de los atributos, elegidos al azar en cada ocasión.

Una vez que tenemos muchos árboles, 1000 por ejemplo, la fase de clasificación se lleva a cabo de la siguiente forma:

- Cada árbol se evalúa de forma independiente y la predicción del bosque será la media de los 1000 árboles. La proporción de árboles que toman una misma respuesta se interpreta como la probabilidad de la misma (Quantdare, 2015, párr. 5).

En esta investigación se utiliza la regresión logística, modelo que es usado para resolver problemas de clasificación, ya que modeliza una probabilidad (probabilidad de fraude); en este caso puede aportar a la identificación del perfil de los empleados que tienen mayor tendencia a cometer fraude interno, lo que servirá para realizar actividades preventivas con estos funcionarios y que los recursos económicos y humanos de la entidad financiera destinados para la investigación se enfoquen en analizar unos riesgos prioritarios que permitan mitigar la posible comisión de un acto de fraude interno.

Se conoce que la regresión logística genera facilidades para estimar e interpretar los parámetros establecidos, además de que entrega la probabilidad de que una nueva observación pertenezca a uno de los grupos (Vélez y Hernández, 2015). Sobre este modelo, IBM (2016) afirma que tiene gran utilidad cuando se busca como objetivo predecir la presencia o, por el contrario, la ausencia de una característica o resultado según unos valores determinados. Adicionalmente, señala que la regresión logística “es similar a un modelo de regresión lineal pero está adaptado para modelos en los que la variable dependiente es dicotómica. Los coeficientes de regresión logística pueden utilizarse para estimar la odds ratio de cada variable independiente del modelo. La regresión logística se puede aplicar a un rango más amplio de situaciones de investigación que el análisis discriminante” (2016, párr. 1).

Por otro lado, también se conoce el CRISP-DM (Cross Industry Standard Process for Data Mining), un método que se enfoca en orientar los trabajos de minería de datos. Dentro de sus fases se incluye la comprensión del negocio y de los datos; la preparación de los datos; el modelado, la

evaluación y la distribución; además de unas tareas necesarias en cada fase y una explicación de las relaciones entre éstas. Dentro de las características de este modelo se resaltan las oportunidades de personalización, la flexibilidad y la creación de un modelo de minería de datos según las necesidades de cada empresa (IBM, 2012).

4. Aspectos metodológicos

Este estudio es una investigación empírica de carácter mixto, con un diseño no experimental (en el sentido de que no se manipulan determinadas variables para observar cómo afectan a otras), exploratorio y descriptivo (pues se trata de un tema poco estudiado aún, que requiere estudios preliminares para ir ahondando en resultados cada vez más precisos y analíticos), que busca ayudar a caracterizar al defraudador interno en una entidad financiera a través del enfoque cuantitativo, el cual “usa la recolección de datos para probar hipótesis, con base en la medición numérica y el análisis estadístico, para establecer patrones de comportamiento y probar teorías” (Hernández, Baptista y Fernández, 2012).

Por otro lado, desde el enfoque cualitativo se aplicó la revisión de literatura como instrumento de investigación para revisar el perfil global del defraudador, a partir de estudios realizados por firmas de auditoría reconocidas mundialmente, como KPMG, PwC y Kroll, y poder así definir las características para identificar a empleados asociados a eventos de fraude interno en la entidad financiera estudiada. Estos objetivos permitieron acceder a información científica sobre el perfil del defraudador en cuanto a la edad, la antigüedad, el cargo, el tipo de fraude, las causas, los intereses, las motivaciones, las estrategias, los impactos, la detección, las técnicas y métodos aplicados para el análisis y las variables de evaluación, datos que contribuyeron al logro del objetivo general de la investigación.

En el caso específico de las variables, estas son importantes al momento de identificar los aspectos a evaluar para obtener un perfil confiable del defraudador. A pesar de que cada empresa tiene características diferentes, al igual que sus trabajadores, existen factores clave que no pueden ser ignorados, como los de tipo demográfico, socio-económico, laboral y transaccional, debido a

que permiten una mejor comprensión de la problemática y orientan el ejercicio de la prevención, control y monitoreo. Así mismo, permiten crear indicadores de la misma naturaleza de las variables e intentar diferenciar los eventos de fraude y no fraude.

Es por esto que las variables utilizadas fueron de tipo demográfico, socio-económico, laboral y transaccional, a saber, cargo, contrato, género, región, jornada, antigüedad, salario y número de alertas en pruebas internas de la entidad financiera. Estas variables se definieron principalmente a partir de la revisión de los perfiles globales del defraudador realizados por KPMG, PwC y Kroll, además de la experticia de los investigadores de la entidad financiera analizada, quienes poseen elementos de criterio para elegir el margen que se acopla de la manera adecuada con el objetivo planteado. Por lo anterior, se espera que las variables seleccionadas contribuyan a que se realice un ejercicio riguroso de discriminación de factores importantes a tener en cuenta y analizar, tanto de la empresa como de los trabajadores, para contribuir a elaborar con un nivel alto de confiabilidad el perfil de riesgo del defraudador interno, y que de esta forma se pueda priorizar el seguimiento a los trabajadores que tienen tendencia a cometer algún fraude interno.

Muchas de las anteriores variables se basan en enorme información transaccional, la cual contiene números bastante elevados respecto a variables de tipo demográfico o de frecuencia, razón por la cual se decidió estandarizar todas las variables numéricas en una escala de [0, 1] para lograr que sean comparables entre sí y poder capturar la mayor variabilidad posible.

Por razones de privacidad y protección de estudios internos de la entidad financiera, se enmascaró el detalle de las variables, pero en la Tabla 1 se da una pequeña descripción acerca del ámbito que fue tenido en cuenta en la construcción de cada una de ellas.

Tabla 1. Medición de las variables dicotómicas y categóricas

<i>VARIABLE</i>	<i>NIVELES</i>	<i>NRO NIVELES</i>
<i>AREA</i>	1= ADMINISTRATIVA, 2= SUCURSAL, 3=TECNOLOGIA	3
<i>CONTRATO</i>	1=DEFINIDO, 2=INDEFINIDO	2
<i>JORNADA</i>	1=TIEMPO COMPLETO, 2=MEDIO TIEMPO	2
<i>REG_NAC</i>	1=CENTRO, 2=BOG y SAB, 3=ANTIOQUIA, 4=SUR, 5=CARIBE, 6=EXTERIOR	6
<i>REG_TRN</i>	1=CENTRO, 2=BOG y SAB, 3=ANTIOQUIA, 4=SUR, 5=CARIBE	5
<i>GENERO</i>	1=MASCULINO, 2=FEMENINO	2
<i>FRAUDE</i>	1=FRAUDE, 2=NO FRAUDE	2
<i>ALERTAS</i>	1=0 ALERTAS, 2=1 ALERTA, 3=2 o más ALERTAS	3
<i>CARGOS</i>	1=JEFE, 2=OPERATIVO, 3=PROFESIONAL	3

Fuente: construcción propia

El instrumento de recolección de información fue la fuente secundaria, a partir de la revisión de literatura en bases de datos bibliográficas de donde se extrajo y analizó el perfil global del defraudador por las diferentes empresas; adicional a esto, por medio de las bases de datos internas de la entidad financiera, se trabajó con la información de los empleados directos vinculados a enero de 2018, al igual que con la base de datos de los empleados desvinculados con justa causa en el periodo de enero de 2011 a junio 30 de 2017, ya que solo a partir de 2011 la entidad financiera empezó a registrar estos datos, que fueron necesarios para la obtención de variables que aportaron en la construcción del perfil de riesgo del defraudador.

Una vez se recopilaron las variables a través de la revisión de literatura y de las bases de datos de empleados de la entidad financiera, se analizaron las variables que incidían en el riesgo de fraude, se utilizaron técnicas estadísticas de predicción, como el modelo Random Forest, para la clasificación de los fraudes, de acuerdo a una probabilidad de ocurrencia. Para estos métodos analíticos se planteó un conjunto de entrenamiento en una proporción del 75% de la información

disponible y uno de prueba, que corresponde a la información restante que se utilizó para la validación del conjunto de entrenamiento; además, se recurrió a medidas de ajuste y comparación entre los modelos conseguidos, para definir cuál es el que mejor se ajusta para llegar al objetivo.

5. Resultados

5.1 Perfil global del defraudador a partir de estudios realizados por firmas de auditoría reconocidas mundialmente

Tomando como referencia el análisis de la encuesta *Perfiles Globales del Defraudador*, realizada por KPMG Internacional (2016), algunos de los hallazgos indican que los defraudadores típicos:

- Tienen entre 36 y 55 años (69 % de los defraudadores investigados).
- Son predominantemente masculinos (79 %), con una proporción cada vez mayor de mujeres, que asciende al 17 % en la actualidad, contra un 13 % en 2010.
- Las mujeres tienden a ser staff o mandos intermedios; los hombres suelen ser directivos.
- Son una amenaza desde dentro (el 65 % son empleados de la empresa).
- Poseen cargos ejecutivos o directivos (35 %).
- El 44 % de los perpetradores tienen autoridad ilimitada en su empresa y son capaces de anular los controles.
- Empleados de la organización con seis años o más de antigüedad (38 %).
- Son descritos como autocráticos (18 %), y tienen 3 veces más probabilidades de ser considerados como amigables que de no serlo.
- Son estimados, se describen a sí mismos como muy respetados en la organización (38%).
- Los controles internos débiles fueron un factor que contribuyó al 61 % de los fraudes.

- Solo el 38% de los defraudadores actúa solo, debido a que el fraude tiene casi el doble de probabilidad de perpetrarse en grupos que en la soledad, puesto que los defraudadores necesitan coludir para eludir los controles.
- Es probable que hayan actuado en conjunto con otros (62 % de los fraudes, un poco más bajo en comparación al 70 % de la encuesta en 2013).
- Están motivados por la ganancia personal y avaricia (60 %) y el sentido de “porque yo puedo” (27 %).
- Los defraudadores masculinos tienden a coludir más que las mujeres, por tanto, también tienden a poseer cargos más altos que las mujeres en la organización.
- Casi una cuarta parte de los estafadores depende de la tecnología.
- Los autores del fraude tecnológico tienden a ser más jóvenes (60 % tienen entre 26 y 45 años de edad).

En el caso del Informe Global de Fraude 2015/2016, de Kroll, se encontró que:

- Más de la tercera parte de los fraudes (36%) fueron cometidos por un miembro del nivel gerencial medio de las compañías afectadas. El 45 % de los fraudes los cometieron empleados junior y el 23 % los cometieron agentes o intermediarios.
- En Colombia el 83% de los encuestados experimentó fraude en su compañía, con un 27% de robo o pérdida de información, o ataque informático. Sigue el conflicto de intereses en la gerencia, con un 23%; un 17 % para el robo de activos físicos o inventario y un 13 % para el fraude de proveedores/proceso de compras.

- Un 44 % de los encuestados colombianos señaló que habían estado involucrados ejecutivos sénior o la gerencia media en los casos de fraude.

Por otro lado, en la Encuesta Delitos Económicos PwC Capítulo Colombia (2016), en síntesis, se muestran los siguientes hallazgos en el país:

- Los principales cargos donde se presentan los defraudadores son los inferiores con un 61%, seguido por los intermedios con un 35%, mientras en las cifras globales los más propensos son los mandos medios.
- La edad de los defraudadores está entre los 31 y 40 años (igual al promedio en Latinoamérica).
- El tiempo de antigüedad en la empresa se encuentra entre los 6 y 10 años, mientras que en América Latina está entre los 3 y 5 años (ver Gráfico 2).
- Las causas de la presencia del fraude son: la oportunidad o habilidad que se tenga para realizar el crimen, con un 86%; la presión que exista para ejecutarlo, señalado por el 9%; mientras que la racionalización fue seleccionada por el 5%.

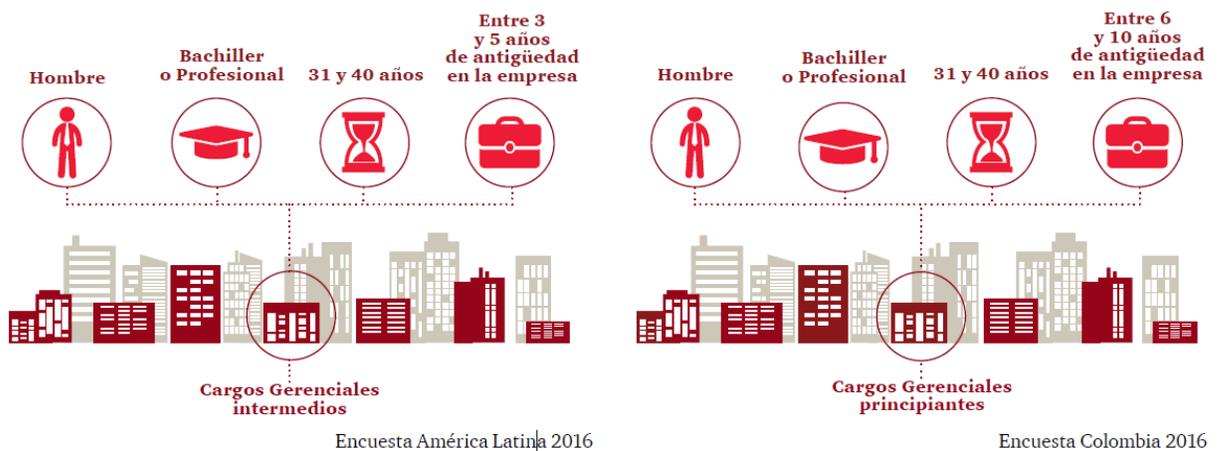


Gráfico 2. Encuesta América Latina Vs Colombia 2016

Fuente: PwC. (2016). Encuesta de Delitos Económicos.

A nivel mundial existen diferentes aportes científicos sobre el defraudador que permiten comprender mejor su perfil, tener en cuenta las principales variables a analizar, los métodos y técnicas, además de poder entender más las causas, consecuencias y detección del fraude. No obstante, estos resultados, presentados para el año 2016 por empresas reconocidas a nivel mundial como KPMG, Kroll y PwC, fueron obtenidos con un enfoque mundial y con una muestra y metodología que no se detallan, a pesar de ser importantes, ya que factores como el tipo de empresa, tamaño, razón social, número de trabajadores; además de las características demográficas, sociográficas y psicográficas de los empleados influyen en los resultados; y, por supuesto, el sesgo de los encuestados, debido al nivel de sus cargos, o el sesgo de la propia encuesta en el caso de KPMG, al elegir un cierto tipo de defraudadores (quizás, los casos más visibles).

Es por eso que estas empresas obtuvieron resultados diferentes en algunas variables. Es, entonces, ahí donde toma mayor relevancia la afirmación de KPMG sobre la necesidad de que cada empresa determine el propio perfil del defraudador según las características, cargos y empleados. Las encuestas, indudablemente, sirven para guiar a las organizaciones en la formulación de las variables que pueden incluir para determinar sus propios perfiles del defraudador, pero es tarea de cada una hallarlos y aprender a controlar el riesgo de fraude.

De acuerdo con la información de las encuestas, puede determinarse que, a nivel global, las características predominantes del defraudador son las siguientes: pertenece al género masculino, puede tener entre más de 30 años y un poco más de 50, puede ocupar cargos altos o medios, lleva entre 6 y 10 años en la empresa, tiene autoridad y capacidad de eludir controles en su empresa, goza de aprecio y respeto, tiende a coludir, usa la tecnología y entre más la usa más joven es, se guía por la avaricia y la ganancia personal y actúa al percibir que tiene oportunidad (dada por la

falta de controles, pero también por el conocimiento adquirido acerca de los procesos de la empresa, debido a la permanencia en ella durante 6 o más años). En Colombia, las diferencias en cuanto al perfil se ubican en tres variables: la edad, que es menor, entre los 30 y 40 años; el cargo, que también es menor, correspondiente a niveles inferiores; y el hecho de que la *oportunidad o habilidad para cometer el crimen* es la causa del fraude en un porcentaje del 86%, por encima de Latinoamérica, con 72%, y el mundo, con 69%. Si bien faltan elementos de análisis, puede decirse que el factor cultural en un país que presenta altos niveles de corrupción y que ha vivido el flagelo de la cultura mafiosa es propicio para que los individuos justifiquen sus conductas antiéticas y vean solo por su interés individual.

Cabe aclarar que cada encuesta se enfoca hacia un tema diferente. Y que solo las de KPMG tienen el perfil del defraudador como asunto relevante. Por eso, algunas aportan datos adecuados para determinar el perfil y otras no. En esta medida, no se utilizó aquí la Encuesta de EY 2016, pues su temática es las conductas corporativas indebidas, llamando la atención sobre el riesgo de fraude que generan conductas de los ejecutivos, que pueden pasar como aceptables. En este sentido, sería recomendable que cada encuesta dedicara una porción importante y coherente con las demás para abordar el tema del perfil del defraudador y construir así un perfil global más ajustado a la realidad.

5.2 Características de empleados asociados a eventos de fraude interno en la entidad financiera

5.2.1 Diagrama de la frecuencia de fraude por cargo

Los datos analizados entre enero del 2011 y el 30 de junio de 2017 arrojaron que 2.548 empleados estaban registrados en la base de infidelidad de la entidad financiera y fueron

investigados por alertas recibidas. 988 casos corresponden a registros de malas prácticas y/o fraude, y los restantes, es decir, 1.560 casos, fueron etiquetados como no fraude.

Dado que en la entidad financiera se identificaron 20 cargos entre los empleados que han incurrido en fraudes y malas prácticas, para producir una mejor asociación de las características en la fase de modelación, se agrupó la variable cargos en 3 niveles: jefe, operativo y profesional, los cuales absorben la mayor parte de la variabilidad de los cargos, que se especifican en el Gráfico 3.

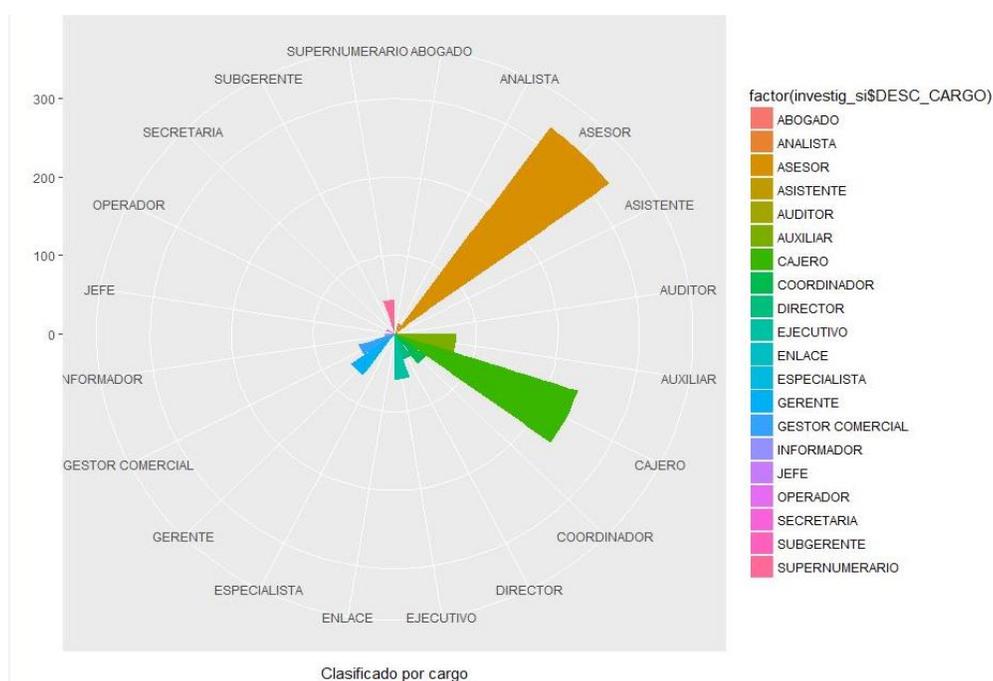


Gráfico 3. Clasificación por cargo
Fuente: Construcción propia

En la Tabla 2 se desglosan los cargos que integran el mando operativo, profesional y directivo, y el número de profesionales que fueron investigados por haber cometido fraude.

Tabla 2. Clasificación por cargo

Operativo (754)		Profesional (129)		Jefe (105)	
Asesor	325	Abogado	1	Director	35
Auxiliar	76	Analista	14	Gerente	66
Cajero	236	Asistente	4	Jefe	2
Enlace	2	Auditor	2	Subgerente	2
Gestor comercial	46	Coordinador	48		
Informador	12	Ejecutivo	59		
Operador	11	Especialista	1		
Secretaria	2				
Supernumerario	44				

Fuente: Construcción propia

Del total de estos trabajadores investigados, el 76%, que representa a 754 personas, pertenece al cargo operativo; seguido por el 13%, que representa a 129 individuos que son profesionales; y el otro 11%, a 105 personas que son jefes (ver Gráfico 4).

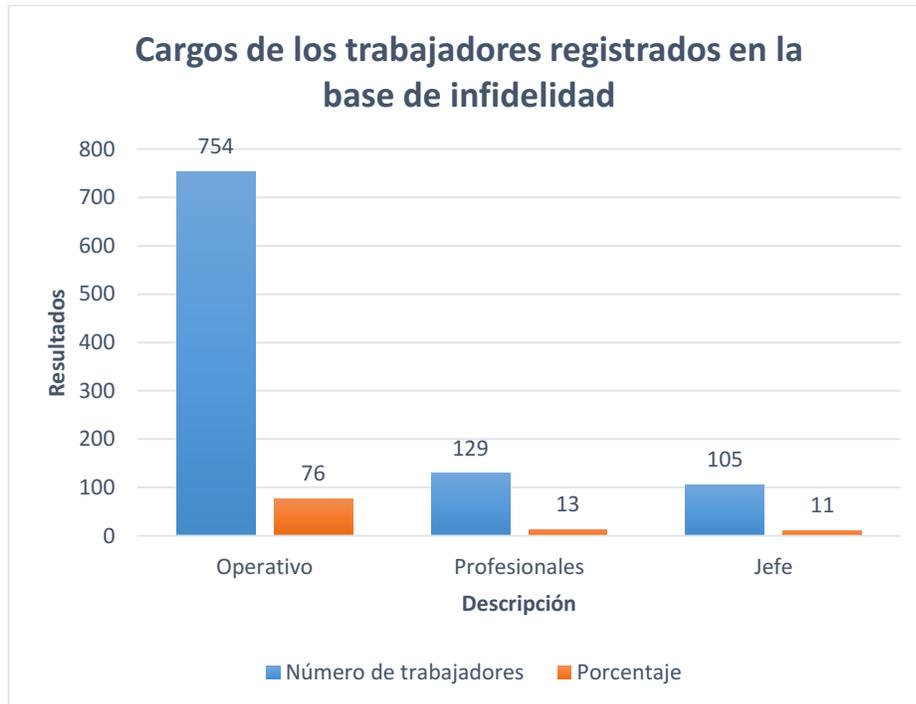
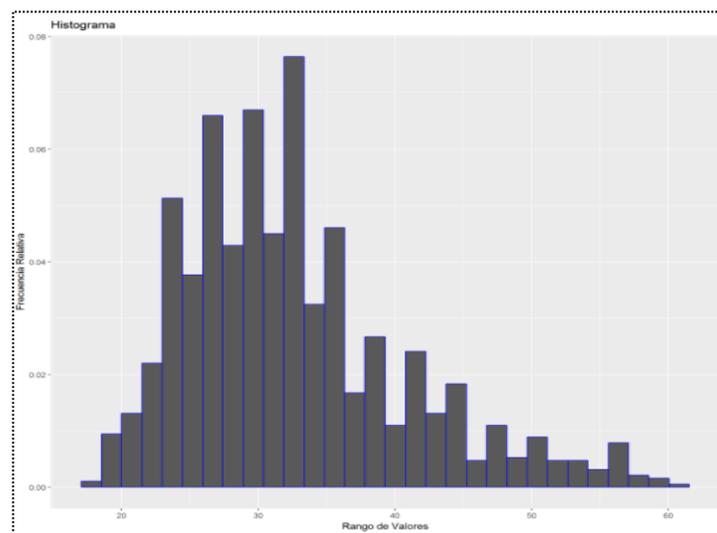


Gráfico 4. Cargos de los trabajadores registrados en la base de infidelidad

Fuente: Construcción propia

5.2.2 Histograma sobre la edad de los funcionarios defraudadores

Después de tener identificados los 988 trabajadores registrados en la base de infidelidad de la empresa objeto de estudio, entre enero del 2011 al 30 de junio de 2017, por incurrir en fraude o mala práctica, se procedió a realizar un histograma de la edad de los empleados defraudadores (o que realizaron malas prácticas), con el objetivo de identificar cómo estaban distribuidas las edades, dividiendo la muestra en cuatro partes (ver Gráfico 5).



	Q1	Q2	Q3 (Percentiles)
##	25%	50%	75%
##	27	31	36

Gráfico 5. Histograma de la edad de los funcionarios defraudadores
Fuente: Construcción propia

Valor Mínimo: 18

Valor Máximo: 64

El primer cuartil o percentil 25, que corresponde al 25% de los empleados defraudadores, poseen una edad menor o igual a 27 años, (227 empleados entre [18-27] años).

El segundo cuartil o percentil 50, que corresponde al 50% de los empleados defraudadores,

poseen una edad menor o igual a 31 años, (233 empleados entre [28-31] años).

El tercer cuartil o percentil 75, que corresponde al 75% de los empleados defraudadores, poseen una edad menor o igual 36 años, (278 empleados entre [32-36] años).

El último cuartil contiene un intervalo entre [37-64] años, con 250 empleados.

Se puede evidenciar con este análisis, entonces, que la edad promedio de los funcionarios defraudadores oscila entre los 27 y los 36 años.

5.2.3 Años de antigüedad del funcionario en la entidad financiera al momento de cometer el fraude o mala práctica

Se realizó un histograma con el tiempo de antigüedad de los empleados defraudadores, con el objetivo de identificar cuánto tiempo llevaban en la empresa al momento de los hechos, dividiendo la muestra en cuatro partes (ver Gráfico 6).

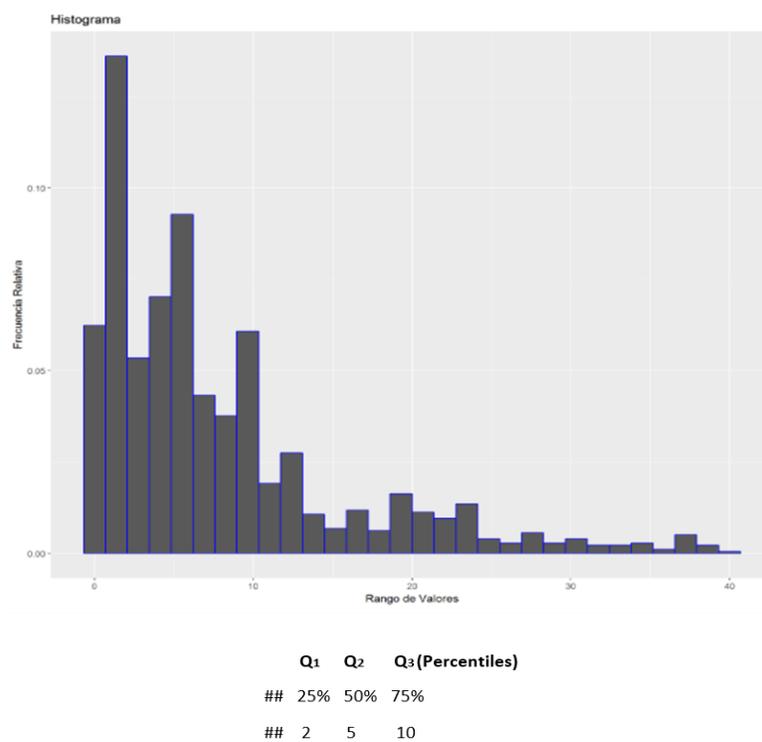


Gráfico 6. Histograma de la antigüedad de los funcionarios defraudadores
Fuente: Construcción propia

Valor Mínimo: 2

Valor Máximo: 10

El primer cuartil o percentil 25, que corresponde al 25% de los empleados defraudadores, poseen una antigüedad menor o igual a 2 años, (314 empleados entre [0-2] años)

El segundo cuartil o percentil 50, que corresponde al 50% de los empleados defraudadores, poseen una antigüedad menor o igual a 5 años, (215 empleados entre [3-5] años)

El tercer cuartil o percentil 75, que corresponde al 75% de los empleados defraudadores, poseen una antigüedad menor o igual a 10 años, (233 empleados entre [6-10] años)

El último cuartil contiene un intervalo entre [11-40] años, con 226 empleados.

Se puede evidenciar que el tiempo de antigüedad del defraudador en la empresa oscila entre 2 y 10 años.

5.3 Evaluación comparativa de dos modelos estadísticos predictivos sobre el perfil de riesgo del defraudador interno en la entidad

5.3.1 Metodología para la evaluación del modelo predictivo

Existen muchos tipos de metodologías para la construcción de modelos propuestos para el desarrollo de proyectos de minería de datos, tales como SEMMA (Sample, Explore, Modify, Model, Assess), KDD (Knowledge Discovery in Databases), DMAMC (Definir, Medir, Analizar, Mejorar, Controlar), CRISP-DM (Cross Industry Standard Process for Data Mining). En este estudio se eligió CRISP-DM, teniendo en cuenta que es un modelo flexible y que se puede personalizar fácilmente; así mismo, porque se ha convertido en la metodología más utilizada,

según un estudio publicado en el año 2007 por la comunidad KDnuggets (Data Mining Community's Top Resource, 2007).

5.3.1.1 Modelos predictivos

En el momento en que se tienen consolidadas las variables de tipo demográfico, socio-económico, laboral y transaccional, se afronta de forma directa el objetivo general del proyecto, el cual es hacer una contribución para definir el perfil de riesgo del defraudador interno en la entidad financiera objeto de estudio. Para definir cómo llevarlo a cabo se probaron dos técnicas estadísticas en el modelado: Random Forest (Bosques Aleatorios) y Regresión Logística, en donde se obtuvo para ambas una medida de ajuste que sirvió para compararlas y decidir cuál era la que acercaba más al objetivo propuesto.

5.3.1.2 Procedimiento para la validación

Para la validación de los modelos que se pusieron a prueba se utilizó la técnica denominada análisis ROC (Receiver Operating Characteristic), la cual grafica la sensibilidad (verdaderos positivos) frente a la especificidad (tasa de falsos positivos), para varios valores de una prueba de diagnóstico (Ramírez *et al.*, 2007).

Un parámetro para evaluar la bondad de una prueba diagnóstica que produce resultados continuos es el área bajo la curva (AUC): una medida para puntuar el rendimiento o efectividad de un modelo estadístico, pues es la probabilidad de que el modelo clasifique una instancia correctamente. El clasificador perfecto tiene una curva que pasa por el punto a una tasa de 100 % de verdadero positivo y 0 % de tasa de falso positivo; cuanto más cerca esté la curva del clasificador perfecto, mejor será para identificar valores positivos. Esta medida se considera aceptable/aceptable cuando se encuentra entre el 0.7 y 0.8 y excelente/bueno cuando clasifique entre 0.8 y 0.9 (Lantz, 2017).

Definiendo un evento a partir de P etiquetas positivas y N negativas, los cuatro posibles resultados se expresaron en una tabla de contingencia o denominada Matriz de confusión, para nombrar correctamente los fallos y aciertos determinantes para calcular el AUC (VER Tabla 6):

Tabla 3. Matriz de confusión

True class → Hypothesized class V	Pos	Neg
Yes	TP	FP
No	FN	TN
	$P=TP+FN$	$N=FP+TN$

Fuente: Japkowicz, N. y Shah M. (2011) Evaluating Learning Algorithms: A Classification Perspective.

5.3.1.3 Procedimiento para la fragmentación de los datos

Teniendo en cuenta que las fases de la metodología CRISP-DM no son tan rígidas, y a medida que se abarca cada fase se hacen otras tareas en segundo plano, partiendo de la preparación de los datos se tomaron en cuenta para la fragmentación los 2.548 registros de la base de infidelidad (de los cuales cabe recordar que 988 corresponden a registros de malas prácticas y/o fraude, y los restantes 1.560 corresponden a casos etiquetados como NO fraude), y se procedió de la siguiente forma dividiendo los datos en dos conjuntos, el de aprendizaje y el de prueba:

- **Aprendizaje (entrenamiento)**

Este conjunto de datos se creó teniendo en cuenta los pesos en un muestreo probabilístico estratificado a nivel laboral de los empleados que han incurrido en algún tipo de fraude y/o malas

prácticas y variables de toda la base original; dicha información corresponde en este caso al 75% de los registros de la base de infidelidad, de la cual 741 casos representan registros de malas prácticas y/o fraude y 1.170, casos de NO fraude, aprendizaje que se usó para los dos modelos de predicción que se utilizaron.

- **Validación (prueba)**

Este conjunto de datos (25% de los registros de la base de infidelidad) quedó formado de igual manera, manteniendo la estructura bajo muestreo probabilístico y garantizando así la aleatoriedad, ya que el modelo desconoce totalmente la información correcta (247 registros de malas prácticas y/o fraude y 390 de NO fraude). De este modo pueden probarse los modelos y así definir cuál es el mejor, es decir, cuál de ellos comete un menor error en la estimación.

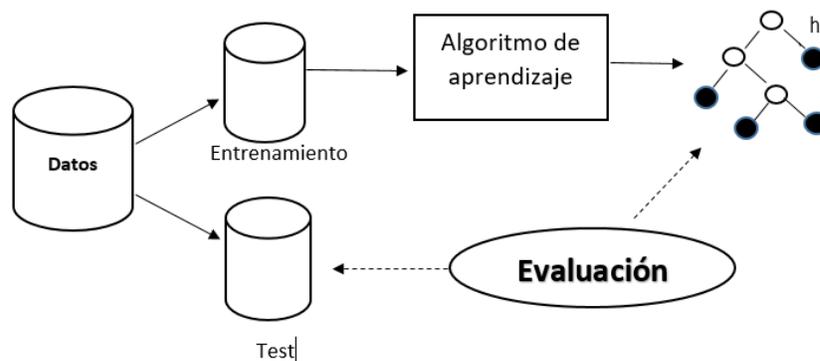


Gráfico 7. Validación

Fuente: Ramírez *et al.* (2007). Introducción a la Minería de Datos

5.3.1.4 Validación Cruzada en Minería de datos

La validación cruzada se usó después de crear una estructura de minería de datos y los modelos de minería de datos relacionados para determinar la validez del modelo, partiendo de crear particiones y hacer un remuestreo en el conjunto de entrenamiento.

En este método estándar de análisis estadístico que resulta muy útil a la hora de desarrollar y

ajustar modelos de minería de datos, los datos se dividieron aleatoriamente en K grupos. Un grupo se reservó para el conjunto de prueba y con los otros K-1 restantes se construyó un modelo para predecir el resultado de los datos del grupo reservado (entrenamiento). Este proceso se repite K veces dejando cada vez un grupo diferente para la prueba. Esto significa que se calculan k ratios de error independientes. Finalmente se construye un modelo con todos los datos y se obtienen sus ratios de error y precisión, promediando los K ratios de error disponibles (Ramírez *et al.*, 2007, p. 36).

Para el objeto de estudio en entrenamiento se realizaron 10 remuestreos (10 k-folds) en la base de fraudes, repitiéndose tres (03) veces para un total de 30 remuestreos.

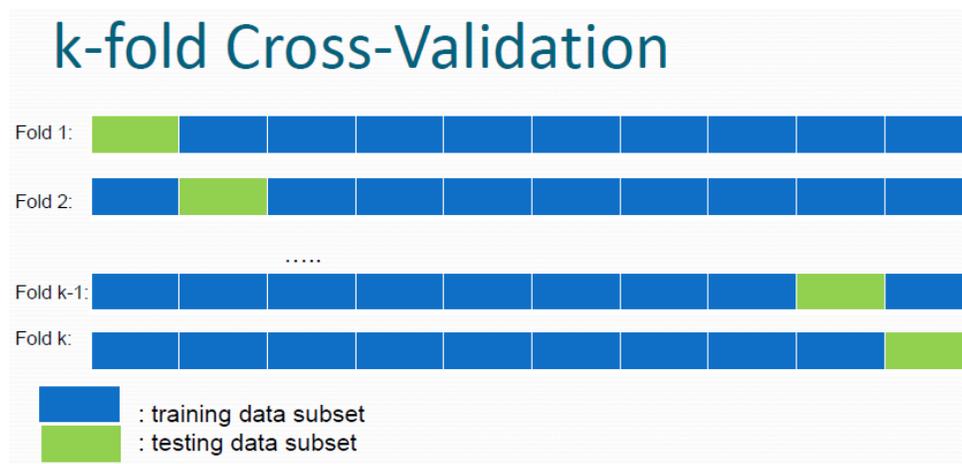


Gráfico 8. Validación cruzada usando k grupos

Fuente Japkowicz, N. y Shah M. (2011). Evaluating Learning Algorithms: A Classification Perspective.

5.3.2 Modelo de Regresión Logística

El modelo de regresión logística se construyó en gran parte de manera empírica, en una serie de intentos en los que se llevaron a cabo varias transformaciones de los datos, dejando como resultado final una transformación en escala [0,1] de todas las variables numéricas y donde se creó una malla de valores para los diferentes parámetros de ajuste, calculando el AUC para cada modelo

creado y seleccionando el mejor.

Teniendo en cuenta el gran número de variables y que la regresión logística en este caso es binomial, es decir, la variable FRAUDE como tal esta etiquetada como 0 o 1, se optó por llevar a cabo una regresión logística penalizada “Red Elástica”, en la cual se penaliza el crecimiento abrupto de los parámetros y se evita así la sobre-parametrización, reduciendo el error de estimación y permitiendo que las probabilidades arrojadas posteriormente tengan más sentido.

Por definición, la regresión logística está dada por:

$$\text{Logit}(\pi) = \ln \left(\frac{\pi}{1 - \pi} \right) = \beta_0 + \beta_1 x_{1,i} + \dots + \beta_k x_{k,i}$$

k = número de variables

i = número de registros

Donde el modelo tiene una formulación equivalente dada por:

$$\pi = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_{1,i} + \dots + \beta_k x_{k,i})}}$$

Luego, para la penalización se define un estimador de red elástica para cada $\hat{\beta}$, como el que minimiza:

$$\sum_{i=1}^k \left(y_i - \sum_{j=1}^k \beta_j x_{i,j} \right)^2 + \lambda_1 \sum_{j=1}^k |\beta_j| + \lambda_2 \sum_{j=1}^k \beta_j^2$$

Donde λ_1 y λ_2 son ambos parámetros de complejidad.

Esta generalización permite imponer mayores penalizaciones sobre variables poco importantes y pequeñas penalizaciones sobre las más relevantes.

De las 30 interacciones realizadas con la validación cruzada del modelo logístico, para las primeras 10 se obtuvieron los siguientes resultados:

Tabla 4. Resultados de la validación cruzada del modelo de regresión logística

	ROC	Sens	Spec	Resample
1	0.6682837	0.7777778	0.3378378	Fold03.Rep2
2	0.6748672	0.8888889	0.4189189	Fold08.Rep1
3	0.6864172	0.8376068	0.3918919	Fold07.Rep1
4	0.6499191	0.8803419	0.3243243	Fold04.Rep2
5	0.7150612	0.8376068	0.3378378	Fold09.Rep2
6	0.7289212	0.8803419	0.3918919	Fold02.Rep1
7	0.6181566	0.7179487	0.3378378	Fold01.Rep1
8	0.6521137	0.8461538	0.3243243	Fold09.Rep1
9	0.6447216	0.8461538	0.3378378	Fold03.Rep1
10	0.6168861	0.8290598	0.3513514	Fold01.Rep3

Fuente: construcción propia

5.3.2.1 Validación

A partir de los resultados anteriores, se planteó la matriz de confusión para relacionar los FP y FN (Falsos positivos y negativos) y las clasificaciones acertadas, dando como foco de importancia la clasificación o efectividad del SI FRAUDE:

Tabla 5. Matriz de confusión

Real →	NO	SI
Predicción ↓		
NO	358	198
SI	32	49

Fuente: construcción propia

Efectividad: 64%

Intervalo de Confianza al 95%: (0.6003, 0.6763)

Sensibilidad: 20%

Especificidad: 91%

- La probabilidad de clasificar de forma correcta a un empleado como no defraudador es del 0,91 (especificidad).
- La probabilidad de clasificar de forma correcta a un empleado como defraudador es del 0,2 (sensibilidad).

5.3.2.2 AUC

Se obtuvo un área bajo la curva del 64%, la cual nos indica que la capacidad predictiva del modelo actualmente es medianamente baja y contiene bastantes falsos negativos, pues se espera llegar a un umbral de AUC cercano al 90% para tener resultados bastante efectivos.

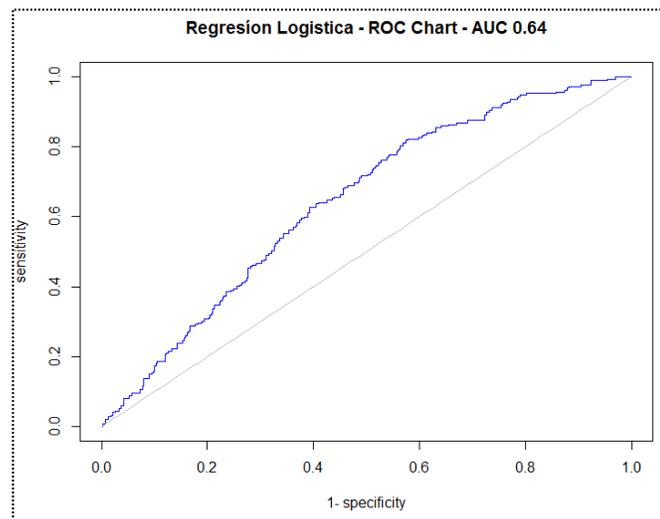


Gráfico 9. Regresión logística – ROC Chart – AUC 0.64

Fuente: Construcción propia a partir del modelo de Regresión Logística

En este caso observamos que un área bajo la curva para este modelo queda en 0.64, lo que se tendrá en cuenta para escoger del modelo a utilizar.

5.3.3 Algoritmo Random Forest

El modelo de *Random Forest* o bosques aleatorios se construyó también en gran parte de manera empírica, con una transformación en escala $[0,1]$ de todas las variables numéricas y donde también se recreó una malla de valores para los diferentes parámetros de ajuste.

El *Random Forest* comienza de manera estándar con un “árbol de decisiones” que, en cuanto al conjunto, corresponde a un aprendizaje. De manera similar que en un árbol de decisión, una entrada inicial se introduce en la parte superior y hacia abajo, y a medida que atraviesa el árbol de los datos, se acumula en conjuntos más y más pequeños; luego, se repite tantas veces como un parámetro de profundidad de árboles lo establezca, como se ejemplifica en el Gráfico 10.

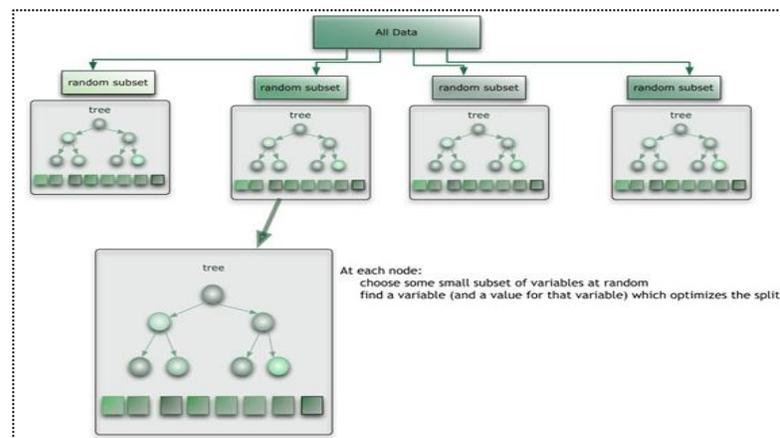


Gráfico 10. All data

Fuente: Randomforest. (2013). All data.

Se llevaron a cabo varias corridas de este modelo de bosque aleatorio. Una de las características que tiene este clasificador es poder visualizar de forma gráfica la importancia de las variables que más aportan a la explicación de la fluctuación de la variable FRAUDE.

Del modelo resultante, las variables importantes, como las demográficas, socio-económicas, laborales y transaccionales (entre estas últimas destacan el número de veces que un empleado ha tranzado con otro empleado que ya ha incurrido o no en eventos de fraude o malas prácticas y

la relación entre salarios y descuentos, pues resultaron ser relevante para el modelo y los resultados obtenidos) o, como tal, variables en las que se encuentra una diferencia en patrones para los defraudadores, basadas en el criterio de Promedio Gini, fueron las siguientes, si bien se presentan enmascaradas por confidencialidad de la información de la entidad financiera:

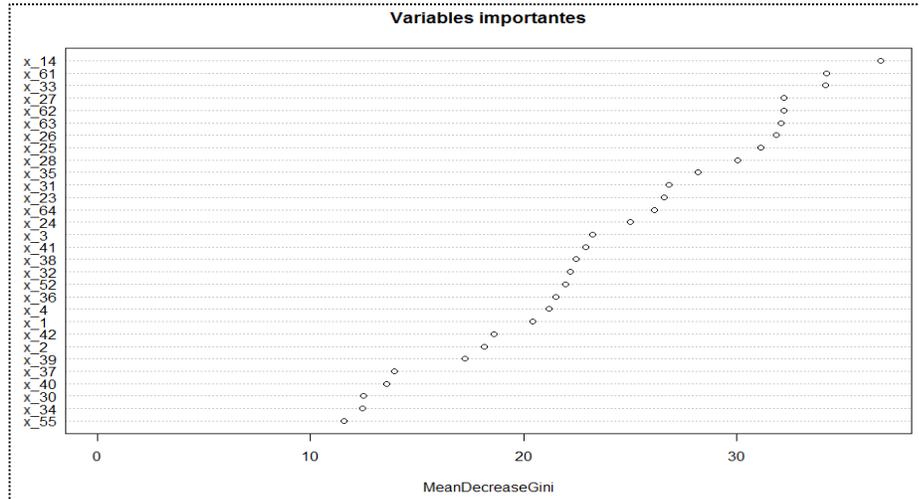


Gráfico 11. Variables importantes enmascaradas

Fuente: Construcción propia

5.3.3.1 Validación

A partir de los resultados anteriores, planteamos la matriz de confusión para relacionar los FP y FN (Falsos positivos y negativos) y las clasificaciones acertadas, dando como foco de importancia la clasificación o efectividad del SI FRAUDE:

Tabla 6. Matriz de confusión

Real →	NO	SI
Predicción ↓		
NO	316	174
SI	74	73

Fuente: construcción propia

Efectividad: 61%

Intervalo de Confianza al 95%: (0.5716, 0.6487)

Sensibilidad: 29%

Especificidad: 81%

- La probabilidad de clasificar de forma correcta a un empleado como no defraudador es del 0,81 (especificidad).
- La probabilidad de clasificar de forma correcta a un empleado como defraudador es del 0,29 (sensibilidad).

5.3.3.2 AUC

Se obtuvo un área bajo la curva del 67%, la cual nos indica que la capacidad predictiva del modelo por ahora es medianamente baja (un poco más alta que el modelo anterior) y presenta bastantes falsos negativos, pues se espera alcanzar un umbral de AUC cercano al 90% para tener resultados bastante efectivos.

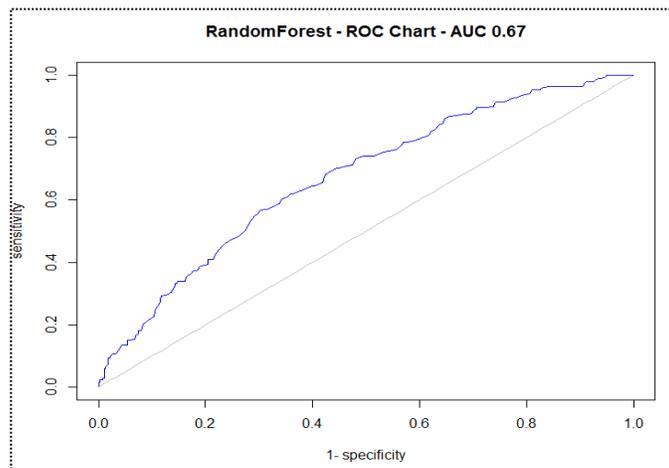


Gráfico 12. Random Forest – ROC Chart – AUC 0.67
Fuente: Construcción propia a partir del modelo Random Forest

Como se puede observar en la matriz de confusión, la precisión del modelo se encuentra en el 0.67%.

5.3.4 Comparación de modelos y análisis

Para la comparación de los modelos, se tuvo en cuenta, además del AUC y efectividad generadas por la matriz de confusión, varias métricas, también a conveniencia relacionadas con el objetivo del proyecto. Se optó por calcular la efectividad de que el modelo clasificara los verdaderos SI fraudes del conjunto de validación, los falsos positivos (por razones de costos de investigación), la efectividad de las alertas generadas para comprobar el porcentaje de aciertos respecto a las alertas o registros clasificados como fraudes.

Este análisis se realizó para toda la línea histórica de los datos de la base de infidelidad desde 2011 hasta 2017 para comprobar de forma general cómo el funcionamiento del modelo clasificaba si se ponía a prueba en los meses en los cuales ya se habían identificado eventos de fraude y/o malas prácticas.

También se hizo el paralelo de las alertas generadas por cada modelo, los aciertos y los fraudes que se identificaron en la base de datos de infidelidad.

Tabla 7. Comparación de modelos

<i>Items</i>	<i>Regresión Logística</i>	<i>Random Forest</i>
<i>Alertas generadas (FP+TP)</i>	81	147
<i>Verdaderos positivos (TP)</i>	49	73
<i>Tasa efectividad TP= (TP/P)</i>	0.198	0.295
<i>Falsos positivos (FP)</i>	32	74
<i>Tasa falsos positivos FP= (FP/N)</i>	0,082	0,189
<i>Verdaderos fraudes P = (FN+TP)</i>	247	247
<i>Efectividad alertada (TP/TP+FP)</i>	0.604	0.496
<i>AUC</i>	0.64	0.67

Fuente: construcción propia

Las categorías analizadas en el anterior paralelo se explican a continuación:

Alertas generadas (FP+TP): el total de alertas generadas se puede identificar a partir de la suma de los falsos positivos (FP), más los verdaderos positivos (TP).

Verdaderos positivos (TP): son las alertas que el modelo clasificó como SI fraude y en la base de infidelidad están marcadas como fraude.

Falsos positivos (FP): son las alertas que el modelo clasificó como SI fraude, pero en la base de infidelidad están marcadas como NO fraude.

Verdaderos fraudes $P = (FN+TP)$: son todos los casos que en la base de infidelidad están identificado como fraudes; en la matriz de confusión se pueden calcular mediante la suma $FN+TP$.

Falsos negativos FN: son las alertas que el modelo clasificó como NO fraude y en la base de infidelidad están marcadas como fraude.

Tasa efectividad $TP = (TP/P)$: es la proporción del total del fraude que el modelo logró identificar como fraude positivo.

Verdaderos positivos TP: son las alertas que el modelo clasificó como SI fraude y en la base de infidelidad están marcadas como fraude.

Tasa falsos positivos $FP = (FP/N)$: es la proporción del total del fraude que el modelo logró identificar como NO fraude, en donde $N=TN+FP$.

Verdaderos negativos TN: son las alertas que el modelo clasificó como NO fraude y en la base de infidelidad están marcadas como NO fraude.

Efectividad alertada: es la proporción del total de alertas; indica cuántas realmente son fraudes y se calcula así: $TP/TP+FP$.

Teniendo en cuenta la información recolectada por las métricas y comparaciones, vemos que el modelo Random Forest clasifica un tanto más por encima (0.295) la “efectividad sí” que el modelo

de regresión logística (0.198), a pesar de que ambos tienen un AUC similar (Medio – Bajo). Específicamente, en la regresión logística se encuentran características como una tasa menor de falsos positivos (1.65) y un mayor porcentaje de clasificación de la “efectividad alertada” (0.604), datos que sustentan la elección de este modelo como el adecuado para alcanzar el objetivo propuesto.

Sin embargo, ya sea con la regresión logística o el *Random Forest*, debe tenerse presente que el evento de fraude interno es, por mucho, más esporádico que otros eventos como fraude externo; y por experticia y análisis descriptivo de los registros de fraude interno que se tenía a disposición, se llegó a determinar que los empleados que cometen fraude monetario y no monetario tienen comportamientos bastante similares en las variables tenidas en cuenta para el modelo. Por el contrario, hay diferencias marcadas en variables como la X-14, la cual hace referencia al SALARIO, dado que la mayoría de cargos críticos o que están más expuestos a cometer actos fraudulentos hacen parte de áreas operativas (las cuales difieren por mucho en salario de las áreas administrativas).

A continuación, se ilustra en el gráfico 13 una simulación de la generalidad que se tiene en los registros, mostrando cómo muchas de las características tenidas en cuenta en la base de infidelidad son similares para ambos eventos (Fraude y/o malas prácticas (988 registros) y NO Fraude (1560 registros)).

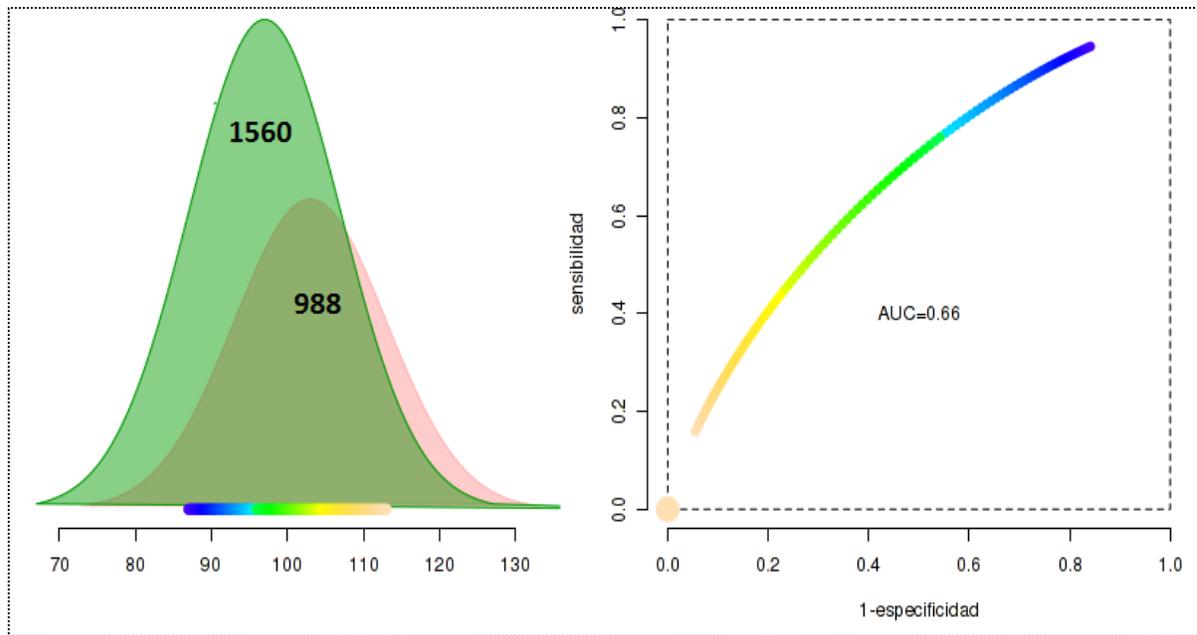


Gráfico 13. Simulación de la generalidad de los actos fraudulentos

Fuente: Construcción propia

Esta complejidad de características también explica la mediana efectividad del modelo (AUC), pues para este es difícil separar patrones o particularidades netas que solo se encuentren en los casos de fraude (región rosa fuera de la intersección).

- El modelo Random Forest nos permitió reconocer el 30% de los casos de fraude y malas prácticas, el cual es un porcentaje aceptable – medio, pues de antemano, por experticia de los investigadores, se conoce que la identificación del defraudador interno no se relaciona solo con las variables demográficas, socio-económicas, laborales y de información transaccional, pues el fraude es un evento inoportuno que está también bastante asociado a las conductas antropológicas y psicológicas del individuo.

5.4 Aplicación del modelo Random Forest sobre la población de la entidad financiera para contribuir a crear su perfil de riesgo del defraudador interno

5.4.1 Características del perfil

A partir de la construcción del modelo de Random Forest, y con el objetivo de encontrar empleados con altas probabilidades de incurrir en actos incorrectos, se realizó una corrida con toda la población activa de la entidad financiera (20.210 empleados al 01 de enero de 2018) para replicar los comportamientos y patrones de los defraudadores en la población. A continuación, se presenta la descripción de aquellos empleados que el modelo predice como defraudadores:

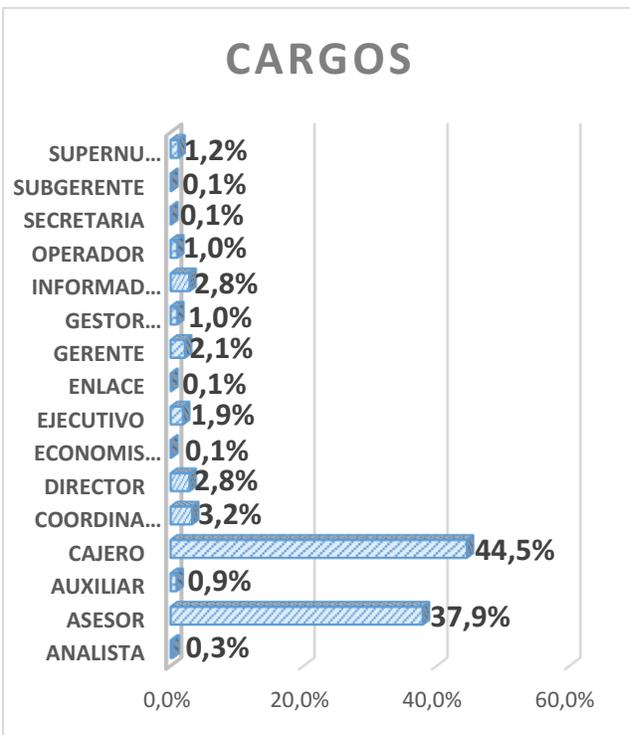


Gráfico 14. Cargos
Fuente: construcción propia

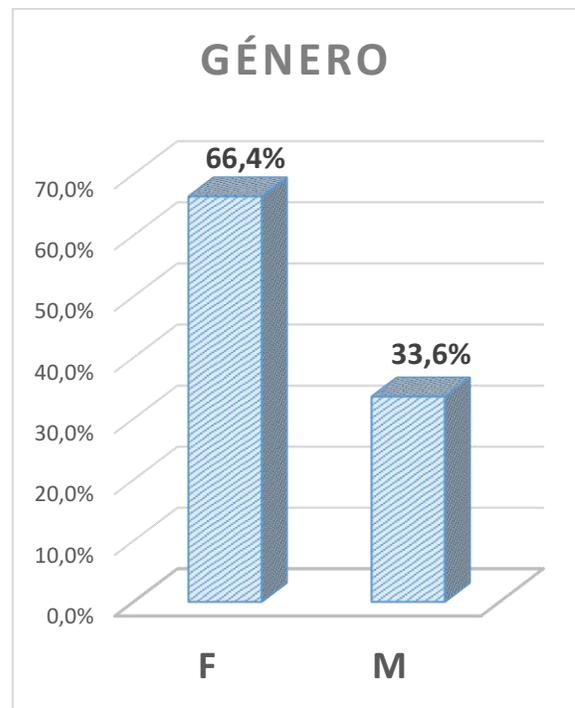


Gráfico 15. Género
Fuente: construcción propia

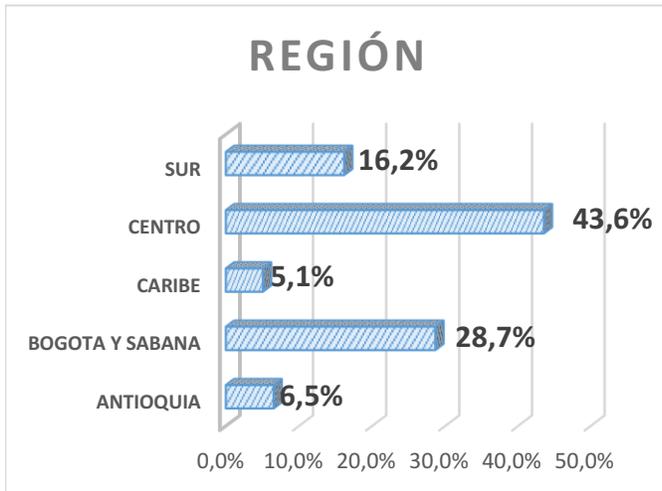


Gráfico 16. Región
Fuente: construcción propia

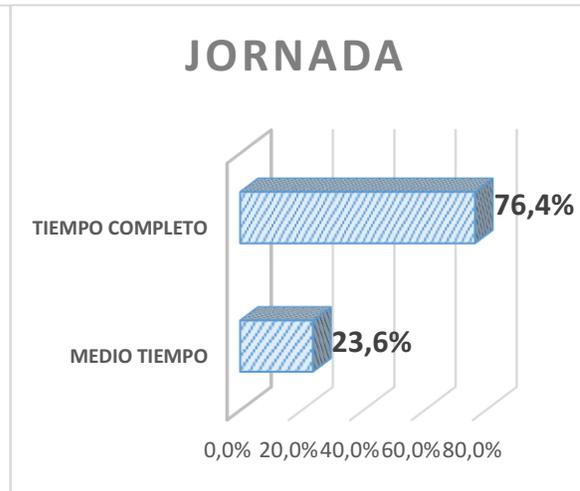


Gráfico 17. Jornada
Fuente: construcción propia

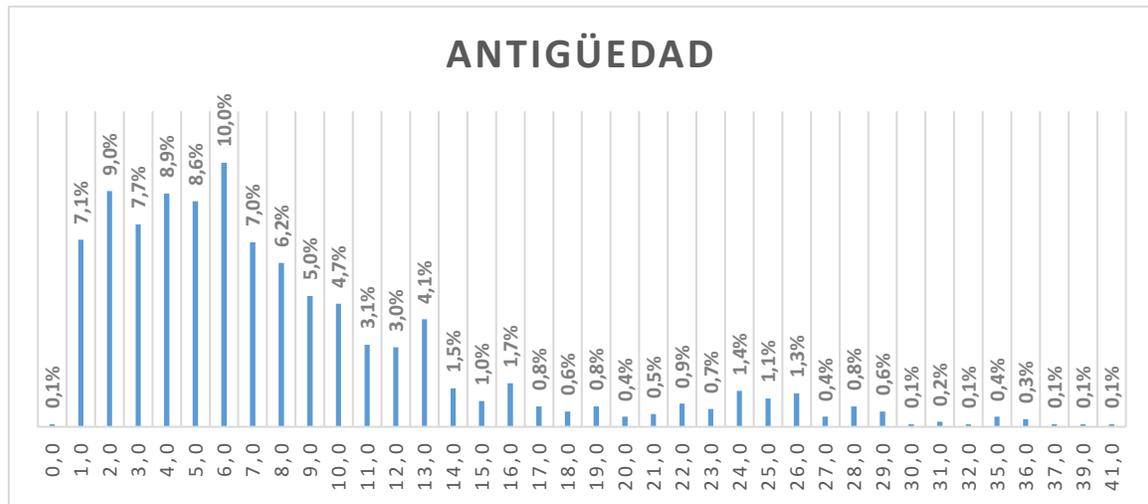


Gráfico 18. Antigüedad
Fuente: construcción propia

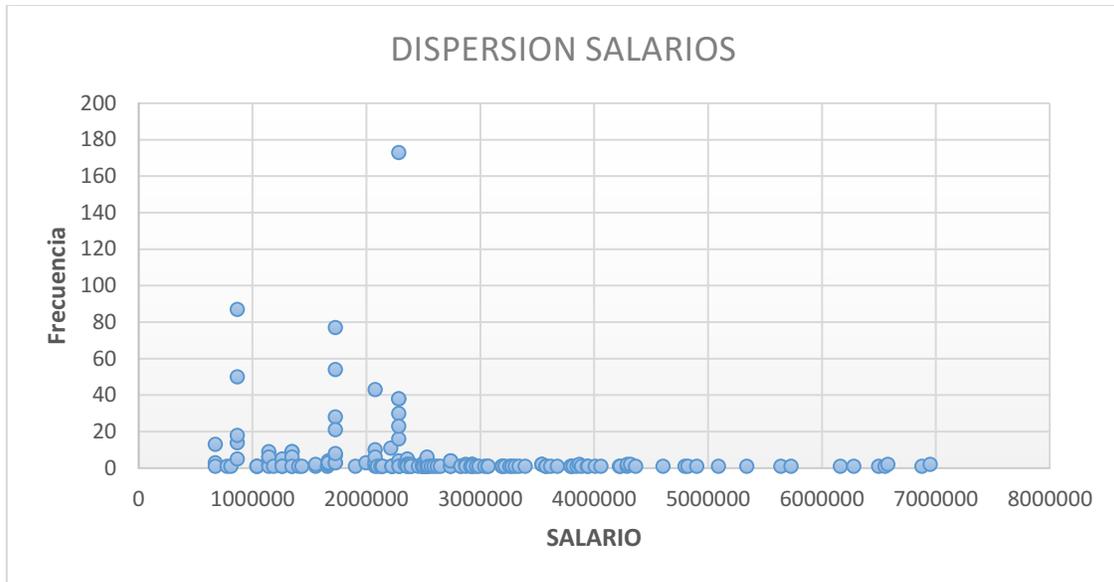


Gráfico 19. Dispersión de salarios
Fuente: construcción propia

5.4.2 Obtención del perfil de riesgo del defraudador mediante un árbol de decisión

Finalmente, teniendo en cuenta que el objetivo de la investigación es construir un perfil parcial del defraudador interno, se entrenó un árbol de decisión utilizando las predicciones y las variables seleccionadas como más importantes por el Random Forest, no con el propósito de obtener otro modelo predictivo, sino una descripción de las predicciones del Random Forest basada en reglas inferidas a partir de la estructura del árbol resultante, reglas que serán útiles para describir el perfil de los empleados con probabilidad de incurrir en actos fraudulentos. Lo anterior es necesario hacerlo, debido a que el modelo resultante del entrenamiento con el Random Forest, por su naturaleza compleja, no permite una fácil interpretación; por tal motivo, no es posible extraer fácilmente reglas de decisión a partir del mismo para describir un perfil de riesgo del defraudador.

Un árbol de decisión es una serie de decisiones o condiciones organizadas en forma jerárquica, a modo de árbol. Encima del nodo raíz de árbol está el problema a resolver. Cada arco conduce a otro nodo de decisión o a un nodo hoja. Los nodos hoja representan la predicción (o clase) del

problema para todas aquellas instancias que alcanza esa hoja. Para clasificar una instancia desconocida, se recorre el árbol de arriba hacia abajo, de acuerdo con los valores de los atributos probados en cada nodo, y cuando se llega a una hoja, la instancia se clasifica con la clase indicada por la hoja (Ramírez *et al.*, 2007, p. 31). Ver el resultado del árbol de decisión en el Anexo 1.

Teniendo en cuenta la información descriptiva y las hojas de tonalidad roja (que reflejan la importancia a nivel probabilístico acerca de las características de un posible defraudador), se extrae la siguiente regla, siendo los más probables los primeros perfiles:

- Empleados que laboren en sucursal bancaria, con ingresos de nómina mayores a \$ 18.000.000.00 millones en los seis últimos meses, que posean alertas en pruebas de minería y con transacciones de retiro mayores o iguales a \$ 4.300.000.00
- Empleados que laboren en sucursal bancaria, con ingresos de nómina mayores a \$ 18.000.000.00 millones en los seis últimos meses, que posean alertas en pruebas de minería, con transacciones de retiro menores a \$ 4.300.000.00, con avances mayores a \$ 350.000.00, con 22 o más transacciones de compras nacionales con tarjeta de crédito y con egresos menores al 7.3% con respecto a lo que reciben diferente de nómina.
- Empleados que laboren en sucursal bancaria, con ingresos de nómina mayores a \$ 18.000.000.00 millones en los seis últimos meses, que no posean alertas en pruebas de minería, con porcentajes de descuentos mayores a 38 % con respecto al ingreso por nómina, con transacciones de retiro mayores o iguales a \$ 4.700.000.00, con 302 o más días sin salir a disfrutar de vacaciones, con un salario mayor o igual a \$ 2.200.000.00 y con ingresos adicionales mayores o iguales a 160 % con respecto a lo que reciben de nómina.
- Empleados que laboren en sucursal bancaria, con ingresos de nómina mayores a \$ 18.000.000.00 millones en los seis últimos meses, que no posean alertas en pruebas de

minería, con porcentajes de descuentos mayores a 38 % con respecto al ingreso por nómina y con transacciones de retiro mayores a \$ 4.700.000.00, con un porcentaje del 0.85 % de ingresos adicionales con respecto a lo que reciben de nómina, con 302 o más días sin salir a disfrutar de vacaciones, con salarios menores a \$ 2.200.000.00, con ingresos por transferencia mayores a \$ 850.000.00 y egresos mayores o iguales a \$ 2.800.000.00.

- Empleados que laboren en sucursal bancaria, con ingresos de nómina mayores a \$ 18.000.000.00 millones en los seis últimos meses, que no posean alertas en pruebas de minería, con porcentajes de descuento menores al 38% con respecto al ingreso por nómina, con más de 18 transacciones de avances con tarjeta de crédito y con transacciones de retiro mayores o iguales a \$ 5.200.000.00.

6. Conclusiones

De acuerdo con la información de las encuestas, a nivel global, las características predominantes del defraudador son las siguientes:

Género: masculino

Edad: entre 30 y poco más de 50 años

Cargos: altos o medios

Antigüedad: entre 6 y 10 años en la empresa

Motivaciones: avaricia, ganancia personal y sentido de la “oportunidad”

Imagen: suelen gozar de aprecio y respeto

Además, suele tener autoridad y capacidad de eludir controles en su empresa, tiende a coludir y usa la tecnología. En cuanto a las causas del fraude, se encuentra la autoridad ilimitada, la falta de controles y la rotación de personal; además, las principales modalidades de fraude son la malversación de activos (fraude interno) y la ciberdelincuencia (que puede corresponder tanto a fraude interno como a externo).

En Colombia, las diferencias en cuanto al perfil se ubican en tres variables:

Edad: entre los 30 y 40 años.

Cargos: niveles inferiores

Motivaciones: *Oportunidad o habilidad para cometer el crimen*. La diferencia en este aspecto radica en que el porcentaje de esta motivación en el país es del 86%, por encima de Latinoamérica, con 72%, y el mundo, con 69%. Hay, entonces, una tendencia más marcada a justificar conductas antiéticas y a velar por el interés individual sobre el colectivo.

Por otro lado, se conoció que los empleados de la entidad financiera objeto de estudio registrados en la base de datos de infidelidad fueron 2.548, de los cuales 988 corresponden a registros de malas prácticas y/o fraude, y los restantes, es decir, 1.560 casos, fueron etiquetados como no fraude. En el caso específico de la cifra correspondiente a las malas prácticas, el 76% (754 personas) pertenece al cargo operativo, seguido por el 13% (129 individuos), que son profesionales, y el otro 11% (105 personas) que son jefes. De los empleados defraudadores, 227 están entre 18-27 años; 233 empleados entre 28-31 años; 278 empleados entre 32-36 años y 250 empleados entre 37 y 64 años. En cuanto a la antigüedad de los empleados defraudadores, 314 llevan entre 0-2 años en la empresa; 215 empleados entre 3-5 años; 233 empleados entre 6-10 años y 226 empleados entre 11 y 40 años.

No obstante, a partir de la aplicación del modelo predictivo de Random Forest, se analizaron variables de tipo demográfico, socio-económico, laboral y transaccional, tales como cargo, contrato, género, región, jornada, antigüedad, salario y número de alertas en pruebas internas de la entidad financiera. Este modelo permitió reconocer el 30% de los casos de fraude y malas prácticas. Este porcentaje no es muy alto, pero no implica que el modelo en sí sea ineficaz, sino que requiere de una base de datos mucho más amplia que aquella de la que se dispuso, ya que la entidad financiera analizada, solo empezó a crear la base de infidelidad a partir del 2011. Así pues, se recomienda a las organizaciones que hagan seguimiento y control a los eventos de fraude y malas prácticas y que creen sus bases de datos lo más completas posibles para que puedan identificar el perfil de riesgo del defraudador característico de su propia entidad. Así mismo, sería recomendable que las organizaciones encontraran métodos para que las áreas de recursos humanos puedan trabajar conjuntamente con las áreas dedicadas a detectar y controlar el fraude, pues los aspectos psicológicos son fundamentales para definir un perfil de riesgo del defraudador mucho

más preciso. Así, la parcialidad del perfil que se definió a través del modelo usado en la investigación, se reduciría si las bases de datos fueran más amplias y se pudiera contar con datos acerca del comportamiento y las motivaciones de los individuos.

Ahora bien, dentro de los resultados arrojados por el modelo, se encontró que los empleados en el cargo de cajero son los que tienen mayores probabilidades de cometer fraude, con un 44.5%, seguido de los asesores, con el 37.9%; las mujeres encajan más en el perfil con un 66.4%, mientras que el 33,6% serían hombres; el 76.4% de los empleados con mayor riesgo de cometer fraude trabaja en jornada completa, mientras que el 23.6% labora medio tiempo; el 43.6% de los trabajadores que son más propensos al fraude son de la zona centro, seguidos por los de Bogotá y la Sabana, con el 28.7%.

Al comparar los resultados del modelo aplicado en la entidad objeto de estudio con las evaluaciones sobre el fraude de las empresas globales de auditoría como KPMG, PwC y Kroll, se evidencia que construir un único perfil del defraudador es un ejercicio que puede ser útil para orientar a las empresas acerca de las variables que pueden analizar para definir sus propios perfiles del defraudador, pues cada empresa, dependiendo de sus actividades, cargos y aún de variables culturales propias de sus países de operación, obtendrá un perfil que puede coincidir en algunos aspectos con los perfiles globales, pero que se diferenciará de un modo u otro.

En el caso de la entidad analizada, es llamativo el alto porcentaje de mujeres que encajan en el perfil de riesgo del defraudador, en oposición a los perfiles globales y para Colombia, pero no es extraño si se analiza que en los cargos críticos de la entidad, el género predominante de los empleados es el femenino. Por otro lado, la coincidencia en el nivel de los cargos en el perfil para Colombia y para la entidad se debe a las características de las actividades que desarrolla: en el cargo de cajero, los empleados tienen más contacto con el efectivo y con el cliente, así como un

número mayor de oportunidades, lo que apoya los resultados de la encuesta de PwC 2016, en cuanto la importancia de la *oportunidad* como causa de la comisión de fraude en Colombia. En ese sentido, una variable de tipo cultural y conductual toma mayor peso y significación.

Ahora bien, los datos arrojados por el modelo pueden ser usados por la entidad financiera para crear un mapa de los cargos críticos y ahondar en características que el modelo, tal y como ha sido aplicado, generaliza. Así, el alto porcentaje de mujeres que encajan en el perfil podría analizarse de acuerdo con el número de mujeres que laboran en la entidad y que se encuentran dentro de los cargos críticos. Un análisis posterior también permitiría definir, por ejemplo, si los cargos más críticos generan pérdidas económicas mayores a las de otros cargos no críticos pero de rango superior y, por tanto, con mayor acceso a la información, a grandes sumas de dinero y con la capacidad de eludir mejor los controles.

A partir de los anteriores hallazgos, se pudo establecer que para la construcción parcial del perfil de riesgo del defraudador interno en la entidad financiera objeto de estudio, el árbol de decisión tuvo en cuenta en una proporción muy alta las variables transaccionales (las más importantes de las cuales fueron las relaciones transaccionales entre empleados y la proporción entre salarios y descuentos) y, por el contrario, no dependió de las variables demográficas, ni de la antigüedad y edad del trabajador. En este perfil se identificaron los empleados que poseen transacciones atípicas y en donde las relaciones de salarios vs los descuentos realizados al empleado no guardan una proporción.

Por otro lado, se logró conocer que el tamaño de los eventos de fraude y/o malas prácticas de la base de datos de infidelidad corresponde aproximadamente al 4.6% de la población total de los empleados de la entidad financiera, y que en su mayoría estos eventos corresponden a malas prácticas.

Así mismo, existe un sesgo natural al tener en cuenta estas proporciones, pues solo el 4.6% de los empleados no puede servir para generalizar de una manera ideal los patrones de fraude de toda la población de la entidad financiera (los posibles cargos, áreas, etc.); sin embargo, la metodología de modelamiento estadístico que se llevó a cabo tiene en cuenta este sesgo natural y permite asignar dichos patrones a individuos que están en la población total de no fraude.

Con el objetivo de mejorar el perfil del defraudador en la entidad financiera, se recomienda para futuras líneas de investigación tener en cuenta las variables que intervienen en el riesgo conductual, la relación de empleados que hayan incurrido en fraude con otros empleados de la entidad, ya que se identificó que el 17 % que corresponde a 170 de los empleados infieles (988) se relaciona en términos transaccionales con otros que han incurrido en actos fraudulentos o en malas prácticas, y los que no han cometido fraude, solo se relacionan con defraudadores en una proporción del 12% que corresponde a 178 empleados de los 1.560 empleados no fraude.

Igualmente, se recomienda incluir, entre otras variables, el análisis de texto de Lync, utilizando la herramienta Watson Personality Insights de IBM, la cual analiza la forma en que las personas escriben para predecir su personalidad, aplicando un análisis lingüístico y de teoría de la personalidad para inferir ciertos atributos a partir de textos desestructurados de la persona.

7. Referencias

- ACFE (2016). Estudio Global de Fraude: Informe a Las Naciones Unidas sobre Fraude y Abuso Ocupacional. Recuperado de: <http://www.acfe.com/rtn2016/docs/2016-report-to-the-nations.pdf>
- ACFE (2017) ¿Qué es el fraude? Recuperado de: <http://www.acfe-spain.com/recursos-contra-fraude/que-es-el-fraude>
- ACFE (2018) Estudio Global 2018: Informe a Las Naciones Unidas sobre Fraude y Abuso Ocupacional. Recuperado de: <https://s3-us-west-2.amazonaws.com/acfe-public/2018-report-to-the-nations.pdf>
- AEECF. (2017) ¿Cuál es el impacto del fraude en las empresas? Recuperado de: https://asociacioncontraelfraude.com/impacto_fraude_empresas/
- ANDI (2017). Régimen bancario. Recuperado de: <http://www.andi.com.co/es/GAI/GuiInv/SisFin/RegBan/Paginas/InsFin.aspx>
- Arcenegui, J. A., Obrero, V. y Martín, J. (2015). Propuesta de un modelo para la prevención y gestión del riesgo de fraude interno por banca paralela en los bancos españoles. *Cuadernos de Contabilidad*, 16(42), 625- 660. Recuperado de: <http://dx.doi.org/10.11144/Javeriana.cc16-42.pmpg>
- Ayala, H. (2008). *El Contador Forense*. Primera edición. Bogotá: Legis.
- Badal V. E. & García C.B. (2016). Detección de fraude financiero mediante redes neuronales de clasificación en un caso real español. *Estudios de Economía Aplicada*, 34(3), 693–709. Recuperado de: <http://search.ebscohost.com/login.aspx?direct=true&db=fua&AN=118462994&lang=es&site=ehost-live>

- Biegelman, M.T. & Bartow, J.T. (2012). Executive Roadmap to Fraud Prevention and Internal Control-Creating A Culture of Compliance. Retrieved from: <https://www.wiley.com/en-us/Executive+Roadmap+to+Fraud+Prevention+and+Internal+Control%3A+Creating+a+Culture+of+Compliance%2C+2nd+Edition-p-9781118004586>
- Bolton, L., Harvey, R., Grawitch, M. y Peluquero, L. (2012). Los comportamientos contraproducentes en el trabajo en respuesta a Agotamiento emocional: un enfoque mediacional Moderado. *Estrés y Salud*, 28: 222 – 233
- Buchahin, M. (2013). Las mejores prácticas para prevención y detección del fraude. *Revista especializada de ACFE – Capítulo México*, (2). Recuperado de: http://www.revistadelfraude.com/Ediciones_anteriores/julio_agosto/Las%20mejores%20practicas%20para%20prevencion%20y%20deteccion%20del%20fraude%20.html
- Calleja, L. (2014). Factores explicativos del fraude empresarial y medidas adoptadas: caso español. Recuperado de: https://addi.ehu.es/bitstream/handle/10810/14550/TFG_Calleja.pdf?sequence=1&isAllowed=y
- Espino, C. (2017). *Análisis predictivo: técnicas y modelos utilizados y aplicaciones del mismo - herramientas Open Source que permiten su uso* (Trabajo de posgrado). Universidad Abierta de Cataluña, Barcelona. Recuperado de: <http://openaccess.uoc.edu/webapps/o2/bitstream/10609/59565/6/caresptimTFG0117mem%C3%B2ria.pdf>
- EY. (2016). 14th Global Fraud Survey, 48. Retrieved from [http://www.ey.com/Publication/vwLUAssets/ey-global-fraud-survey-2016/\\$FILE/ey-global-fraud-survey-final.pdf](http://www.ey.com/Publication/vwLUAssets/ey-global-fraud-survey-2016/$FILE/ey-global-fraud-survey-final.pdf)
- Fida, R., Paciello, M. y Tramontano, C. (2014). Un enfoque integral para entender los

comportamientos contraproducentes en el trabajo: El papel de factores de estrés, las emociones negativas, y la desvinculación moral. *J Ética Bus*, 130: 131-144.

Grijalva, E. y Newman, D. (2015). Narcisismo y comportamiento en el trabajo contraproducente. *La psicología aplicada: una revisión internacional*, 64(1): 93-126.

Hernández Sampieri, R., Baptista Lucio, P., y Fernández Collado, C. (2012). *Metodología de la investigación*. 6a. edición. Estados Unidos: MC GRAW HILL.

IBM. (2012). Manual CRISP-DM de IBM SPSS Modeler. Recuperado de: <ftp://ftp.software.ibm.com/software/analytics/spss/documentation/modeler/15.0/es/CRISP-DM.pdf>

IBM. (2016). Regresión Logística. Recuperado de: https://www.ibm.com/support/knowledgecenter/es/SSLVMB_22.0.0/com.ibm.spss.statistics.help/spss/regression/idh_lreg.htm

Instituto de Contabilidad y Auditoría de Cuentas (ICAC). (2013). *Norma Internacional de Auditoría 240*. Recuperado de: <http://www.icac.meh.es/NIAS/NIA%20240%20p%20def.pdf>

Institute of Internal Auditors, IIA (2013). Normas internacionales para el ejercicio profesional de la auditoría interna. Recuperado de: <https://na.theiia.org/translations/PublicDocuments/IPPF%20Standards%20Markup%20Changes%202013-01%20vs%202011-01%20Spanish.pdf>

Japkowicz, N. y Shah M. (2011) Evaluating Learning Algorithms: A Classification Perspective.

Kassem, R. y Higson, A. (2012). The New Fraud Triangle Model. *Journal of Emerging Trends in Economics and Management Series*, 3(3), 191-195. Recuperado de: <http://jetems.scholarlinkresearch.com/articles/The%20New%20Fraud%20Triangle%20Model.pdf>

- KPMG. (2013). Encuesta de fraude Colombia. Recuperado de: <http://uniajc.edu.co/controlinterno/wp-content/uploads/2015/11/Encuesta-de-Fraude-en-Colombia-2013.pdf>
- KPMG. (2016). Perfiles Globales del defraudador: Presente y futuro de los delitos económicos. Recuperado de: <https://assets.kpmg.com/content/dam/kpmg/pdf/2016/05/profiles-of-the-fraudster.pdf>
- Kroll. (2016). Global fraud & risk report. Recuperado de: <https://www.kroll.com/en-us/intelligence-center/reports/global-fraud-risk-report>
- Lantz, B. (2015). Machine Learning with R - Second Edition. Birmingham: Packt Publishing.
- Logicalis. (2015). Modelos predictivos: reforzando el valor de una buena decisión. Recuperado de: <https://blog.es.logicalis.com/analytics/modelos-predictivos-reforzando-el-valor-de-una-buena-decision>
- López, W. y Sánchez, J. (2011). El Triángulo del Fraude y sus Efectos sobre la Integridad Laboral. *Anales de Estudios Económicos y Empresariales*, 21, 39-57. Recuperado de: <file:///C:/Users/JP/Downloads/Dialnet-ElTrianguloDelFraudeYSusEfectosSobreLaIntegridadLa-4115401.pdf>
- McMahon, R., Pence, D., Bressler, L., & Bressler, M. S. (2016). New Tactics in Fighting Financial Crimes : Moving Beyond the Fraud Triangle. *Journal of Legal and Regulatory Issues*, 19(1), 16–26.
- Mondragón, J. (2011). Análisis Predictivo para detectar el fraude con tarjeta de crédito. Recuperado de: <https://www.ibm.com/developerworks/ssa/local/analytics/prevencion-de-fraude/index.html>
- Muñoz, P. y Álvarez J. A (1997). Comportamiento del Mercado: Hipótesis alternativas. *Revista de*

Bolsas y Mercados Españoles, 60, 29-33.

- Olcina, E. (2016). Motivaciones y perfil del defraudador en la empresa. *Revista de Contabilidad y Dirección*, 3, 11-25. Recuperado de: https://www.economistas.es/Contenido/EC/Motivaciones_y_perfil_del_defraudador_en_la_empresa.pdf
- Olmedo, F., Velasco, F., y Valderas, J. M. (2007). Caracterización no Lineal y Predicción no Paramétrica en el IBEX35. *Estudios de Economía Aplicada*, 25(3), 815–842.
- Omar, A., Vaamonde, J.D. y Delgado, H. (2012). Comportamientos contraproducentes en el trabajo: diseño y validación de una escala. *Revista Diversitas - Perspectivas en Psicología* - Vol. 8 (2): 249-265.
- Penney, L.; Hunter, E. y Perry, S. (2011). El comportamiento de la personalidad y trabajo contraproducente: Usando la conservación de la teoría de los recursos para reducir el perfil de los empleados desviadas. *Revista de Psicología de las Organizaciones y el Trabajo*, 84, 68 – 77.
- Pujari, A. K. (2012). *Data Mining Techniques*. Estados Unidos: Universities Press.
- PwC. (2014). Encuesta Sobre Fraude y Delitos Económico. Recuperado de: <https://www.pwc.es/es/publicaciones/gestion-empresarial/assets/encuesta-fraude-economico-2014.pdf>
- PwC. (2016). Hacia una nueva ética de los negocios: preparados para evitar el crimen económico y cibernético. Recuperado de: <https://www.pwc.com/co/es/publicaciones/crime-survey-2016.pdf>
- Quantdare. (2015). Random forest vs simple tree. Recuperado de: <https://quantdare.com/random-forest-vs-simple-tree/>

- Ramírez, N. D. (2008). El fraude en la actividad bancaria. *El Cuaderno - Escuela de Ciencias Estratégicas*, 2 (4), 279- 296. Recuperado de: <https://dialnet.unirioja.es/descarga/articulo/2991253.pdf>
- Ramírez, M. J., Hernández, J., y Ferri, C. (2007). *Introducción a la Minería de datos*. Madrid: Pearson Educación.
- Random Forest. (2013). All data. Recuperado de: <http://randomforest2013.blogspot.com.co/2013/05/randomforest-definicion-random-forests.html>
- R Notebook. (2016). Ejemplo ML para detección de fraude en TC. Recuperado de: https://rstudio-pubs-static.s3.amazonaws.com/298528_845f68bfb9814f05bade678fc770bc99.html
- Robinson, S. L., & Bennett, R. J. (1995). A typology of deviant workplace behaviors: A multidimensional scaling study. *Academy of Management Journal*, 38(2), 555–572. doi:10.2307/256693.
- Spector, P. E., & Fox, S. (2005). A stressor-emotion model of counterproductive work behavior. In S. Fox & P. E. Spector (Eds.), *Counterproductive work behavior: Investigations of actors and targets* (pp. 151–176). Washington, DC: American Psychological Association.
- Wolfe, D.T. y Hermanson, D. R. (2004): The fraud diamond: considering the four elements of fraud. *The CPA Journal*, December, 1–5. Recuperado de: <https://digitalcommons.kennesaw.edu/cgi/viewcontent.cgi?referer=https://www.google.com.co/&httpsredir=1&article=2546&context=facpubs>