

SEGMENTACIÓN DE CONTRIBUYENTES EN BASE A ANALÍTICA AVANZADA Y DESARROLLO DE ACCIONES PERSONALIZADAS

Eduardo Medel González

MBA, Universidad de Chile

Magister en Dirección y Gestión Tributaria, Universidad Adolfo Ibáñez

Ingeniero Civil Industrial, Universidad de Chile

emedel09@gmail.com

1. INTRODUCCIÓN

Una estrategia de gestión del cumplimiento se orienta al análisis de las obligaciones tributarias desde el punto de vista del riesgo de incumplirlas. El estudio del riesgo permite indagar en los grupos de contribuyentes más propensos a incumplir, así como en las causas asociadas al incumplimiento, las debilidades o fortalezas de los procesos asociados al cumplimiento y la identificación de las diferentes actividades que puede desarrollar la Administración Tributaria para mitigarlas, lo que resulta relevante en el actual contexto de digitalización, automatización y comunicación en base a tecnologías de información, que cada vez más, caracteriza la relación entre autoridades fiscales y contribuyentes.

La gestión de cumplimiento basada en riesgos, desde el punto de vista del análisis de las capacidades tecnológicas, desafía a las Administraciones Tributarias en al menos estas cinco necesidades fundamentales:

- i. Conocimiento del contribuyente: sus características principales, obligaciones tributarias que tienen, comportamientos, disposición al cumplimiento y causas de los incumplimientos. Necesidades, percepciones, entorno y ámbito de negocio, entre otros factores que permiten una mirada holística del contribuyente.

- ii. Información: existencia, calidad, disponibilidad, gobierno y uso efectivo de los datos, que permita transformarlos en conocimiento institucional y apoyar de manera efectiva la toma de decisiones institucional. Para este proceso de transformación, las tecnologías de información juegan un rol fundamental por cuanto el volumen, la velocidad de actualización y la variedad de tipos (imágenes, sonidos, textos) no podrán ser procesadas y analizadas a escala humana.
- iii. Herramientas tecnológicas: orientadas a la explotación de datos (cada vez más integrado con visualización), análisis (procesamiento y modelamiento de datos) y visualización de datos (dashboards interactivos que permitan una mejor lectura de la información, el “data storytelling”, que en definitiva da sentido y permite una mejor interpretación de los datos) Importante consignar que sin estas herramientas tecnológicas no sería posible analizar los datos y convertirlos en información institucional.
- iv. Competencias analíticas: equipos profesionales con las competencias técnicas para el diseño e implementación de modelos analíticos, para “hacer hablar los datos” y lograr transmitir dichos hallazgos al conocimiento institucional. Estas competencias actualmente son escasas no sólo en las Administraciones Tributarias, sino a nivel del mercado en general. Se suele hablar de “data scientist” o experto en ciencia de datos, que viene a ser un profesional con una particular combinación de competencias tales como gestión de datos, estadística, modelamiento y computación, por ponerlo en simple, es un nuevo perfil que ha nacido también con el desarrollo tecnológico y el abanico de posibilidades que abren las nuevas tecnologías disponibles.
- v. Innovación de tratamientos: capacidad para desarrollar diferentes formas para gestionar el cumplimiento de los contribuyentes, que permitan mitigar, eliminar, transferir o aceptar los riesgos de incumplimiento. Buscar soluciones distintas a problemas complejos, hacer eco a la célebre frase de Einstein y hacer cosas distintas para lograr resultados distintos. Fortalecer el uso de las tecnologías para alcanzar mayor certeza tributaria con altos niveles de eficiencia institucional y aprovechar sinergias resultantes de la digitalización de las relaciones, y aplicarlas a la relación de las Administraciones Tributarias con los contribuyentes.

En consideración a las necesidades identificadas y al marco de referencia en el que se desarrolla la actividad de las Administraciones Tributarias, en este documento se buscará profundizar en las técnicas disponibles para gestionar el cumplimiento tributario, focalizando en la necesidad de segmentar a los contribuyentes, analizando

diferentes modelos, ámbitos para hacerlo, y abriendo espacios de discusión en los que explorar o adoptar soluciones que permitan hacer el trabajo más eficiente en el contexto de la información y tecnologías disponibles que desafían el accionar de sector público y privado.

2. ESCENARIO PARA GESTIONAR EL CUMPLIMIENTO

Desde el punto de la Administración Tributaria, la gestión del cumplimiento puede ser clasificado como un desafío de complejidad muy alta, dado que debe considerar múltiples aspectos, entre otros: la cantidad y diversidad de contribuyente (actividades, tamaños, regímenes, volúmenes de operaciones, cantidad de clientes, proveedores y empleados), la dinámica de los negocios y estructuras asociadas (creación y término de empresas, modelos de negocios, modificaciones e integraciones), la diversidad y permanente actualización de las temáticas tributarias (marco legal, normativo, impuestos, franquicias y procesos) y la evolución de las técnicas y tecnologías disponibles (sistemas, softwares y capacidades de procesamientos)

Ahora bien, sin perjuicio de este complejo ecosistema, la Administración Tributaria debe buscar estrategias que permitan maximizar el cumplimiento tributario, asegurando la satisfacción de los contribuyentes, con la restricción de los recursos disponibles. Esta sencilla fórmula de optimización, muy fácil de escribir, por cierto, se transforma en un desafío institucional no trivial, desde el cual nacen las estrategias relacionadas a dividir el problema, “dividir para reinar” atribuido al Príncipe de Maquiavelo y también a Sun Tzu, en el sentido de transformar problemas complejos en múltiples problemas sencillos. De esta forma, se buscará focalizar las acciones institucionales, con lo cual la necesidad de conocer a los contribuyentes pasa a ser imprescindible, siendo la segmentación y el conocimiento de la disposición al cumplimiento, piezas claves de esta focalización, las cuales pueden ser abordadas sobre la base de las herramientas tecnológicas, que vienen a funcionar como facilitadores para simplificar problemas, transformarlos a escala humana, facilitando así la toma de decisiones en escenarios automatizados.

2.1 Volumen y diversidad de contribuyentes

Dada la importancia de conocer al contribuyente y la enorme diversidad de características y contextos del universo de contribuyentes, surge la inevitable necesidad de segmentarlos para poder contar con distintos grupos, que por una parte faciliten el entendimiento de su comportamiento y por otra, haga más manejable la gestión de sus obligaciones.

Como referencia, para dimensionar la diversidad de características de los contribuyentes, consideremos que, de acuerdo con el Servicio de Impuestos Internos¹:

- i. 10,2 millones de contribuyentes personas naturales, impuesto global complementario y único de segunda categoría. 7,6 millones en tramo exento, 1,6 millones en el tramo del 4% y 83 mil personas en el tramo superior de impuestos con una tasa del 35%.
- ii. 1,2 millones de personas jurídicas², de las cuales 1,1% son grandes empresas y representan del orden del 86% de las ventas, 2,2% son empresas medianas y representan del orden de 5,8% de las ventas y 60% son pequeñas y representan del orden del 6,2% de las ventas³.
- iii. 360 mil contribuyentes hicieron inicio de actividades, es decir, nuevos contribuyentes ingresaron al sistema. A su vez 140 mil hicieron término de giro, es decir, abandonan su vida como contribuyentes⁴.
- iv. 3,4 millones de contribuyentes presentaron declaración anual de impuesto a la renta, Formulario 22, y 972 mil estaban obligados a presentar la declaración mensual y pago simultáneo de impuestos, Formulario 29⁵.
- v. Existen del orden de 630 actividades económicas, agrupadas en cerca de 230 subrubros, los que a su vez se agrupan en cerca de 20 rubros, entre estos últimos, Comercio al por mayor y al por menor; Reparación de vehículos automotores y motocicletas (31%); Agricultura, ganadería, silvicultura y pesca (9%); Transporte y almacenamiento (9%); Construcción (8%); Industria manufacturera (7%); Actividades profesionales, científicas y técnicas (6%) y Actividades de alojamiento y de servicio de comidas (5%).

1 Se utiliza el año 2017 para poder comparar las diferentes cifras disponibles en Portal de Estadísticas Servicio de Impuestos Internos: <http://www.sii.cl/destacados/ogp/> y en Gestión y estadísticas del Servicio de Impuestos Internos: http://www.sii.cl/sobre_el_sii/gestion_y_estadisticas.html

2 Empresas Grandes: ventas anuales superiores a 100 mil UF.
Empresas Medianas: ventas anuales entre 25 mil y 100 mil UF.
Empresas Pequeñas: ventas anuales entre 2,4 mil y 25 mil UF.

3 Se debe consignar que, de acuerdo a la página del Servicio de Impuestos Internos citada, existe un no despreciable 16% de contribuyentes sin información de ventas.

4 Importante considerar que este número ha tenido un incremento significativo en cuanto a la serie histórica de términos de giro, lo que se explica por modificaciones introducidas por La Ley N° 20.780 de 2014, a los artículos 68 y 69 del Código Tributario, que inciden en materia de inicio de actividades y término de giro, facultando al Servicio de Impuestos Internos a realizar el término de giro cuando cuente con los antecedentes que permitan establecer que una persona, entidad o agrupación sin personalidad jurídica, ha cesado su giro o actividad económica, sin que haya dado el aviso respectivo.

5 Cuenta Pública 2018 correspondiente a gestión 2017.

2.2 Disposición al cumplimiento y cantidad de obligaciones

Los contribuyentes en su ciclo de negocio enfrentan diversas obligaciones tributarias, desde el registro o inicio de actividades, la entrega de información, las declaraciones y pago periódicos de impuestos y el proceso de término de giro cuando las actividades que se desarrollan llegan a su fin. Dependiendo de los tipos e intensidad de actividades que desarrolla, entre otros factores, le afectará una mayor o menor cantidad de obligaciones. Estas obligaciones estarán determinadas a su vez por los regímenes tributarios, franquicias, tipos de desembolsos e ingresos que se desarrollan como parte de su actividad. Se aumentarán las obligaciones de información, y complejidad de la tributación en la medida se desarrollen negocios en Chile y en el exterior. De igual forma, diferentes obligaciones nacerán según las formas jurídicas que adopten las empresas.

Desde el punto de vista de la disposición a cumplir se presenta uno de los interesantes dilemas a resolver en la gestión de cumplimiento, cual es si se está o no pagando la correcta cantidad de impuestos definido por la legislación e instrucciones de la autoridad tributaria. Aun cuando sea posible determinar esta probabilidad, la causa no es simple de detectar y requiere una indagatoria adicional, por cuanto puede estar asociada a razones desde el desconocimiento del contribuyente por la complejidad del sistema o ausencia de información, hasta el comportamiento poco ético, cuando se erosionan bases tributarias mediante la utilización de esquemas de planificación tributaria agresiva o derechamente se evade el pago de impuestos.

Cualquiera sea el set de obligaciones que debe cumplir un contribuyente siempre encontraremos dos lados en la ecuación, por un lado, la probabilidad que la Administración Tributaria seleccione a un contribuyente para revisión y detecte su incumplimiento (convengamos que subsiste una probabilidad condicional de manera implícita) y por otro, la probabilidad de incumplir del contribuyente. Ambos lados tienen costos y beneficios. No revisar o revisar y no detectar diferencias existiendo estas, implica un costo para la Administración Tributaria. Incumplir y esperar no ser revisado puede ser atractivo para un contribuyente especulador, que aplicando teoría de juegos puede buscar un equilibrio conveniente. Para evitar este tipo de dilemas, cual Dilema del Prisionero, o al menos reducir el costo de la administración y aumentar el costo de la evasión, cobra importancia contar con sistemas sancionatorios eficaces, pero a su vez con estrategias de segmentación que permitan dividir los problemas, fortalecer y focalizar las estrategias preventivas y modelar la disposición al cumplimiento, temas que pueden ser aplicados con las técnicas que iremos profundizando en los apartados siguientes.

3. SEGMENTAR CONTRIBUYENTES

El conocimiento de los contribuyentes permite a la Administración Tributaria analizar diversas características de estos, identificar similitudes y agruparlos en función de dichas características con el propósito de diferenciar su estrategia de gestión de cumplimiento.

De acuerdo con la Real Academia Española (RAE), “segmentar” corresponde a “cortar o partir algo en segmentos”, y “segmento”, en lo que a la materia en análisis se relaciona, corresponde a “cada uno de los grupos homogéneos diferenciados, a los que se dirige la política comercial de una empresa.” De esta forma, para efectos de una Administración Tributaria, un segmento de contribuyentes es un conjunto de contribuyentes que comparten algunas características o atributos, que permiten clasificarlos como “similares” con el propósito de gestionar su cumplimiento tributario de una manera más eficiente.

Los contribuyentes pueden ser segmentados de múltiples formas, lo importante es el sentido de negocio, lo que, aplicado a los objetivos de la Administración Tributaria, significa que un proceso de segmentación debe apoyar el diseño de las estrategias para gestionar el cumplimiento tributario, es decir:

- i. Focalizar los esfuerzos de control, en cuanto al tipo e intensidad de acciones de tratamiento a desplegar:
 - a) Determinar grupos relevantes en cuanto tamaño, y, en consecuencia, impacto tributario en recaudación
 - b) Identificar segmentos de alta complejidad tributaria, que requieren un nivel de competencia mayor de parte de los equipos fiscalizadores
 - c) Identificar segmentos con mayor riesgo de incumplimiento o fraude
- ii. Focalizar herramientas de asistencia o facilitación, en cuanto a su efectiva orientación a las necesidades del contribuyente
 - a) Determinar grupos de contribuyentes que requieren capacitación o asistencia, y en que materias o trámites las requieren
 - b) Identificar procedimientos que deben ser simplificados o customizados,
 - c) Difundir temas o alertas según necesidades

Para efectos de las descripciones y análisis posterior es importante realizar algunas definiciones y alcances, para efectos de simplificar la redacción y entendimiento de las ideas principales:

- i. “U” corresponde al universo de contribuyentes cuyo cumplimiento debe gestionar una Administración Tributaria. Cada contribuyente que compone el universo U puede ser descrito de acuerdo con diversas características.
- ii. Las características son en un sentido amplio, atributos o variables de un contribuyente, algunas compartidas o comunes y otras únicas. Es posible definir una cantidad muy variada de atributos, de diferente naturaleza, pero cuyo objetivo es entregar información sobre el contribuyente que permita describirlo:
 - a) x_{ij} corresponde al atributo j del contribuyente i
 - b) Cada atributo x_{ij} puede ser clasificado de acuerdo con la naturaleza de la característica que describe:
 - Categóricos, aquellos que como su nombre lo indica, representan una categoría, son cualitativos, discretos y los podríamos clasificar en nominales (número de teléfono, domicilio) u ordinales por sentido de orden (tamaño de empresas: pequeñas, medianas y grandes)
 - Numéricos, que representan una cantidad expresada en un número, cuantitativos, pueden ser discretos (cantidad de sucursales) o continuos (ventas, compras o capital)
- iii. X_i corresponde a un vector de atributos del contribuyente “i” que permite agrupar las diferentes características de los contribuyentes, es decir, los diferentes atributos del contribuyente “i”. De esta forma $X_i = (x_{i1}, x_{i2}, x_{i3}, \dots, x_{in})$ es el vector de atributos del contribuyente i, donde “n” es la cantidad de atributos con las que se lo describe.
- iv. $C(K)$ es una partición o segmentación del universo U de contribuyentes, conformada por K segmentos o clústeres, es decir, K grupos de contribuyentes que comparten ciertas características o atributos. Cada partición $C(K)$ requiere necesariamente de un conjunto de atributos que permita evaluar que tan similares o distintos son los contribuyentes entre sí.

Existen diversas formas de particionar el universo U de contribuyentes, en cuanto cantidad de segmentos y forma o lógica de partición, dependiendo de los objetivos y sentido de negocio buscado, es decir, si en definitiva

entrega o no valor a las acciones que debe realizar la Administración Tributaria.

Así, por ejemplo, podemos decir que $C(2)$ es una partición o segmentación del universo U de contribuyentes según tipo de persona, que está conformada por 2 segmentos, por ejemplo, personas naturales y personas jurídicas. Por otro lado, $C(630)$ es una segunda forma de particionar o segmentar el universo U de contribuyentes en 630 segmentos, de acuerdo a la actividad económica principal que estos desarrollan. De acuerdo al registro del SII, según vimos en apartado anterior, se identifican 630 actividades económicas. Cada contribuyente tiene una y sólo una actividad principal por definición. Finalmente, $C(3)$ podría ser una nueva partición del universo U de contribuyentes, en cuanto a la clasificación del riesgo de incumplir obligaciones tributarias, establecida en tres segmentos, por ejemplo, Alto, Medio y Bajo.

- v. S_k corresponde a uno de los K segmentos o clústeres de la partición $C(K)$. Por definición, todos los contribuyentes que pertenecen a S_k son similares entre sí y además son distintos a aquellos contribuyentes que pertenecen a un clúster distinto a S_k . Se dirá que un segmento S_k estará mejor definido mientras mayor sea la similitud de los contribuyentes que pertenecen a él y mientras menor sea la similitud con los contribuyentes que pertenecen a segmentos diferentes a S_k dentro de la misma partición $C(K)$.

Por ejemplo, si consideramos la partición anterior $C(2)$, tendremos que S_1 corresponde a todos los contribuyentes cuyo atributo tipo de persona corresponde a las personas naturales y S_2 a las personas jurídicas. Si consideramos la segmentación que corresponde a $C(630)$, tendríamos por ejemplo que, S_1 agrupa a todos los contribuyentes cuyo atributo de actividad económica es la Construcción, S_2 la Minería, S_3 el Transporte y así sucesivamente hasta S_{630} que podría ser Comercio. Si consideramos $C(3)$, tendríamos que S_1 corresponde a todos los contribuyentes de riesgo Alto, es decir, que representan un riesgo significativo para la Administración Tributaria y sobre los cuales es posible se desplieguen acciones de manera más frecuente e intensa.

- vi. Para determinar la pertenencia a un determinado segmento, se requiere necesariamente de un conjunto de atributos, que permita evaluar que tan similares o distintos son los contribuyentes entre sí, y modelos matemáticos que permitan aplicar reglas y condiciones de clasificación. Usualmente estas reglas, al igual como ocurre en ámbitos sociales “dime con quién andas y te diré quién eres” están basadas en mediciones de distancia. En

este ámbito, existen diversas medidas de distancia, Manhattan, Jaccard, Coseno y Minkowski entre otros, una de uso general, es la distancia Euclidiana, que corresponde a un caso particular de Minkowski y queda determinada de la siguiente forma:

$$D(X_i, X_j) = \sqrt{\sum_{k=1}^n ((x_{ik} - x_{jk}))^2}$$

Donde:

- a. X_i vector de atributos de contribuyente i , tal que $X_i = (x_{i1}, x_{i2}, x_{i3}, \dots, x_{in})$
 - b. X_j vector de atributos de contribuyente j , tal que $X_j = (x_{j1}, x_{j2}, x_{j3}, \dots, x_{jn})$
- vii. Es posible definir que un contribuyente i , cuyo vector de atributos es $X_i = (x_{i1}, x_{i2}, x_{i3}, \dots, x_{in})$, pertenece al segmento S_k , cuando se cumpla al menos una de las siguientes condiciones:
- 1) Todos o algunos de los atributos x_{ij} cumplen una determinada regla.
 - 2) Una combinación de los atributos x_{ij} , cumplen una determinada regla.
- viii. Consideremos un par de ejemplos para graficar lo anterior:

a. Ejemplo 1:

- i. Consideremos al Contribuyente 1 con un vector de atributos $X_1 = (x_{11}, x_{12}, \dots, x_{1n})$ y al Contribuyente 2 con un vector de atributos $X_2 = (x_{21}, x_{22}, \dots, x_{2n})$
- ii. Podemos determinar si pertenecen a un mismo segmento, evaluando la distancia euclidiana entre ambos vectores X_1 y X_2 .

$$D(X_1, X_2) = ((x_{i1} - x_{j1})^2 + \dots + (x_{in} - x_{jn})^2)^{1/2}$$

- iii. Si la distancia entre ambos vectores es cero o menor que un valor arbitrario determinado por un modelo estadístico, será posible considerar a X_1 y X_2 como parte de un mismo segmento.

b. Ejemplo 2:

- iv. Una segmentación del universo de contribuyentes en cuanto tamaño, expresado según volumen de ventas o capital propio tributario, genera 3 segmentos: Grandes, medianas y pequeñas empresas.
- v. Los atributos con los cuales es posible segmentar el universo son $X_i = \{\text{Ventas, Capital propio tributario}\}$ y una regla de asignación a un segmento podría ser $\text{Ventas} < 1.400 \text{ UTM}$ y $\text{Capital} < 1.350 \text{ UTM}$.
- ix. Como se puede desprender de lo analizado y de la experiencia tributaria en general, existen múltiples particiones y segmentos en los cuales es posible agrupar a los contribuyentes, lo cual depende de las características de estos y de los objetivos perseguidos. De esta forma, un contribuyente puede pertenecer a diferentes segmentos de diferentes particiones. La combinación de estos segmentos puede entregar valiosa información a la Administración para efectos de la gestión de cumplimiento. Ejemplo: cantidad de contribuyentes de tamaño grande, de cierto sector económico y con cierto nivel de riesgo de incumplimiento. Esta combinación de segmentos entrega luces a las estrategias que puedan ser desplegadas para cautelar el cumplimiento tributario de manera aún más específica que aquella que solo mira un segmento de una partición. Evidentemente las posibilidades de combinaciones están sujetas a las capacidades analíticas en cuanto a información disponible para el diseño e implementación de atributos, competencias y herramientas analíticas para construir y visualizar los resultados.

4. ANALÍTICA TRIBUTARIA SOBRE SEGMENTOS

Cuando se habla de segmentar se puede tender a encasillar el problema a aquellos modelos de segmentación o clusterización, sin embargo, es importante conocer que diversos modelos estadísticos pueden ser utilizados para segmentar, así como analizar o describir características o comportamientos de un segmento. En este sentido, antes de cualquier referencia a modelos estadísticos, es preciso consignar que más allá de la técnica que se elija, subyacen aspectos fundamentales en la concepción, diseño e implementación de un modelo analítico, cualquiera sea el ámbito que se trate, lo que permite dimensionar su alcance y aplicación al problema que se busca resolver. De esta forma, a modo de preámbulo:

- i. Los modelos analíticos buscan transformar datos en información, descubriendo patrones en los datos, luego, los datos pasan a ser el insumo más importante de este proceso de descubrimiento, el mineral más cotizado para esta minería. Pero cuidado, existe actualmente mucha información disponible, en diferentes formatos, volúmenes y velocidades de acumulación, luego cobra importancia la selección adecuada de esta.
- ii. El desarrollo de un modelo analítico requiere un concienzudo entendimiento del problema que se busca resolver, la definición clara del objetivo perseguido, la rigurosa selección de las técnicas y un arduo trabajo con datos, lo que necesariamente implica contar con las metodologías, competencias técnicas y herramientas tecnológicas adecuadas para hacerlo. Existen tres metodologías muy reconocidas y de amplio uso entre los especialistas en la materia CRISP-DM (Cross-Industry Standard Process for Data Mining), SEMMA (Sample, Explore, Modify, Model, Assess)⁶ y KDD (Knowledge Discovery in/from Databases) la que comprende 5 etapas: Selección, Pre-Procesamiento, Transformación, Modelamiento y Evaluación.
- iii. En la etapa de Modelamiento, encontraremos diversos modelos de minería de datos, los que principalmente pueden ser del tipo predictivo, cuando buscan predecir el valor de una variable a partir de un set de datos, o modelos descriptivos, cuando buscan entregar información adicional a través de la descripción de los datos disponibles. Por otro lado, los modelos también pueden ser clasificados de acuerdo al tipo de aprendizaje, siendo Supervisados, cuando se busca un patrón en los datos a partir de información que conocemos, o No Supervisados, donde no se tiene mayor conocimiento de los datos, es decir, no existen ejemplos a los que el modelo pueda recurrir para descubrir un patrón⁷.

6 Metodologías de análisis de datos.

CRISP-DM (Cross-Industry Standard Process for Data Mining), que propone 6 etapas para el proceso de minería de datos: comprensión del negocio, comprensión de los datos, preparación de los datos, modelado, evaluación e implementación.

KDD (Knowledge Discovery in/from Databases) que consta de 5 etapas: selección de los datos con los que se trabajará, preprocesamiento o de limpieza, transformación, minería de datos, que es donde se obtienen los patrones buscados, y la culmina con la etapa de interpretación y evaluación. SEMMA (Sample, Explore, Modify, Model, Assess) que a su vez también consta de 5 etapas: muestreo de datos, proceso de exploración, modificación, modelamiento y evaluación.

7 Los modelos de aprendizaje supervisados están asociados a problemas de clasificación o regresión, es decir, buscan entregar un patrón con una salida categórica o numérica respectivamente. Árboles de decisión, Regresiones lineales, Regresiones logísticas, Redes neuronales, Redes Bayesianas, Naive Bayes, Support vector machine (SVM) y K-vecinos cercanos (KNN) entre otros, son algoritmos habituales de este tipo.

A continuación, se describirán algunos ejemplos de aplicación de modelos de minería de datos a problemáticas usuales de la gestión de cumplimiento tributario que enfrentan las Administraciones Tributarias:

- i. Agrupar contribuyentes a partir de ciertas características (vector de atributos) para apoyar el diseño y focalización de las estrategias de asistencia y apoyo en el cumplimiento de las obligaciones tributarias:
 - a. A través de un algoritmo de aprendizaje no supervisado se podría indagar la existencia de segmentos a partir de diferentes atributos, por ejemplo, relacionados a frecuencia de contacto con la Administración Tributaria, canales de comunicación utilizados (web, teléfono y oficina), errores o problemas tributarios que tiene, cantidad de obligaciones, tipo de declaraciones y formularios que presentan, tipos y cantidad de códigos de formularios que utilizan.
- ii. Analizar a los contribuyentes a partir de ciertas características para efectos de determinar si el nivel de ingreso declarado a la Administración Tributaria y, en consecuencia, el impuesto determinado, se condicen con los niveles esperados por la Administración Tributaria:
 - a. Modelos del tipo no supervisado podrían agrupar contribuyentes similares según la actividad económica desarrollada, ubicación geográfica, número de empleados, fecha de inicio de actividades, tipo de activos, participaciones societarias y cantidad de sucursales, entre otros atributos. Para cada clúster se podría analizar la distribución de los ingresos e impuestos pagados, de tal forma de encontrar a aquellos contribuyentes cuyos ingresos o impuestos no se encuentren en un rango aceptable (por ejemplo, cuando el valor se encuentre entre el promedio menos dos desviaciones estándares y el promedio más dos desviaciones estándares, o según análisis de cuartiles o quintiles)
 - b. Modelos supervisados de regresión, podrían estimar los ingresos esperados a declarar a partir de un set de atributos, similares a los del punto anterior, lo que permitiría evaluar la distancia entre la estimación y lo declarado. Obviamente es una simplificación dado que los ingresos deberían ser ajustado dado que pueden variar por diferentes motivos tales como, shock externo, tipo de industria, indicadores de actividad económica u otros, los que eventualmente podrían formar parte del modelo e ir aprendiendo en ciclos.

Los modelos no supervisados se orientan a problemas de segmentación, asociaciones, reducción de dimensiones, outliers y correlaciones. K-means, Agrupación jerárquica, DBSCAN (agrupación por densidad), A priori y Fuzzy C-means, son algoritmos habituales de este tipo de aprendizaje.

- iii. Encontrar patrones de contribuyentes falladores de alguna obligación tributaria, por ejemplo, contribuyentes que con mayor probabilidad: no presentarán una declaración jurada o un formulario de declaración de impuestos; no pagarán un giro de impuestos emitido por la Administración Tributaria; emitirán documentos irregulares o presentarán una declaración con inconsistencias:
 - a. Un modelo supervisado de clasificación puede determinar si un contribuyente pertenece a la clase cumplidora o no, a partir de un set de atributos, por ejemplo, historial de cumplimientos, segmento de tamaño, actividad económica, edad tributaria, monto de impuestos, cantidad de documentos, clientes o proveedores, empresas relacionadas, representantes legales u otros atributos. En este caso, se requiere información de los datos en cuanto falladores y no falladores, para que el algoritmo que se utilice pueda, a partir de ejemplos, determinar un patrón en los datos. Un modelo de regresión podría devolver la probabilidad con la que se estima el contribuyente fallará la obligación tributaria cuyo cumplimiento se modela.
- iv. Analizar y clasificar documentos para generar recomendación a contribuyentes, por ejemplo, sugerir instrucciones o procedimientos administrativos, jurisprudencia administrativa o judicial de acuerdo a las características de un contribuyente o búsquedas que este realice en el sitio web de la Administración Tributaria. Por ejemplo, un contribuyente de una determinada actividad económica y régimen tributario puede requerir información sobre gastos deducibles o franquicias a las que tiene derecho, con lo cual un sistema podría mostrar, de manera similar a lo realizado por diferentes plataformas comerciales, las búsquedas que han realizado contribuyentes similares a él e información disponible relacionada a su consulta, pero focalizada a su perfil, respecto de sus particularidades tributarias.
 - a. Un modelo supervisado de clasificación de textos podría permitir analizar las palabras que contienen los diferentes documentos que dispone la Administración Tributaria tales como instrucciones o jurisprudencia. En este caso se requiere la confección de un diccionario de palabras que permita analizar y clasificar los textos. Al tratarse de un modelo supervisado, como se ha mencionado, se requiere una etiqueta, es decir, un documento con este tipo y frecuencias de palabras o leyes citadas que pertenecen a una determinada categoría. Un segundo algoritmo puede asociar documentos a una consulta particular y perfil de contribuyente, generando aprendizaje a partir de la evaluación (aciertos y errores) respecto de las recomendaciones efectuadas.

- v. Encontrar patrones de aquellos contribuyentes que realizarán término de giro, con el propósito de contactarlos de manera temprana y desplegar estrategias de asistencia o control oportuno en la etapa de cese de actividades, la cual conlleva una serie de trámites que pueden tener errores o retardo en su realización.
 - a. Un modelo supervisado de regresión podría entregar un valor entre 0 y 1, equivalente a la probabilidad con la que un contribuyente terminará su giro en un determinado periodo tributario, a partir de la información de términos de giro históricos. Atributos que podrían ser relevantes en este caso: la evolución en el nivel de actividad del contribuyente, expresado como cantidad de clientes, proveedores, cantidad de documentos emitidos o volúmenes de ventas e impuestos pagados, cantidad de empleados, tipos de giros, correlación con actividad económica nacional o internacional (IMACEC; precio de commodities tales como cobre, litio, molibdeno, petróleo, u otros parámetros industriales)
 - b. De manera similar, en el mismo fenómeno, un modelo supervisado de regresión podría estimar la extensión de la vida de un contribuyente a partir de un set de atributos, de tal forma que esa información podría ser de utilidad para programar actividades de monitoreo de la actividad del contribuyente para efectos de preparar el acompañamiento adecuado y oportuno al cierre de actividades.
- vi. Determinar el nivel de cumplimiento tributario futuro de un contribuyente a partir de una mirada holística de su comportamiento tributario y niveles de cumplimiento histórico de las obligaciones tributarias. Un caso particular guarda relación con aquellos contribuyentes que acaban de enrolarse o realizar inicio de actividades, en cuyo caso una dificultad adicional radica en la ausencia de información histórica del contribuyente que permita analizar su comportamiento.
 - a. Una combinación de modelos supervisados de clasificación y regresión podrían determinar diversos componentes del cumplimiento tributario del contribuyente, tales como, la probabilidad de fallar en ciertas obligaciones o ser clasificados en clases cumplidoras o no. En este caso, cada obligación (declarar, informar o pagar) puede ser modelada de manera independiente y los resultados de cada uno de los modelos, puede combinarse para tener la vista integrada del cumplimiento del contribuyente.
 - b. En el caso de nuevos contribuyentes, un posible set de atributos puede estar conformado por aquellos que determinan o están

relacionados con el comportamiento tributario de los socios o representantes legales de los nuevos contribuyentes, si existen.

- vii. Clasificación de textos de facturas para identificar tipos de productos comercializados, con lo cual se puede determinar un patrón de compras y ventas permitiendo validar la actividad económica informada y el correcto uso de créditos fiscales (En el caso del IVA, el derecho a crédito fiscal requiere que las compras estén asociadas al giro o actividad económica que se desarrolla)
 - a. Un modelo supervisado de clasificación de textos, de manera similar al comentado sobre clasificación de documentos, puede permitir determinar los productos contenidos en una factura y determinar el set de productos que pueden estar asociados a una actividad económica. Se requiere un set de datos etiquetados y algunas reglas que permitan asociar y excluir productos de ciertos giros o actividades económicas.

5. EJEMPLOS DE SEGMENTOS Y TÉCNICAS ANALÍTICAS ASOCIADAS

En el contexto tributario existen diversas formas de segmentar a los contribuyentes para efectos de apoyar el análisis de comportamientos y consecuentemente la detección de las necesidades de los contribuyentes y de la Administración Tributaria para gestionar el cumplimiento de las obligaciones tributarias. Este proceso, en definitiva, apoya la definición de las estrategias que se desplegarán, en cuanto a su tipo, alcance, foco e intensidad. Algunos ejemplos de estrategias sobre segmentos definidas en el Plan de Gestión de Cumplimiento Tributario (PGCT) 2019, del Servicio de Impuestos Internos:

- i. Mejorar nivel de cumplimiento de obligaciones tributarias derivadas de la compra y venta de productos y servicios a través de medios digitales o electrónicos.
- ii. Mejorar niveles de cumplimiento de segmento de altos ingresos y patrimonio, respecto de bases imponible y créditos del Impuesto Global Complementario.
- iii. Disminuir la erosión de la base imponible y el traslado de beneficios al exterior por parte de contribuyentes del segmento Multinacionales.

- iv. Facilitar y asegurar el cumplimiento de las obligaciones tributarias del segmento de pequeños contribuyentes, habilitándolos para que dicho cumplimiento sea lo más informado, fluido, correcto y fácil posible.
- v. Reducir brechas de cumplimiento entre otros de los segmentos Minería del cobre, Construcción de viviendas, Tabacos y cigarrillos, Acuicultura, Banca e Industria de bebidas analcohólicas.

Entre las diferentes formas de segmentación de contribuyentes se pueden citar aquellas determinadas por el tamaño del contribuyente, la actividad económica que desarrolla, los impuestos que componen su carga tributaria, los procesos en los que participa o demanda de la Administración Tributaria, o los riesgos de incumplimiento que caracterizan su comportamiento tributario, entre otros tipos de segmentación usualmente encontrados como parte de las estrategias de Administración Tributaria.

5.1 Segmentación según tamaño del contribuyente

5.1.1 Segmentación según reglas del negocio

Los segmentos definidos por tamaño de contribuyentes suelen tener una pequeña variante, por cuanto se trata de una segmentación en dos fases, primero se segmenta el universo de contribuyentes según el tipo de persona (natural o jurídica) y luego las personas jurídicas o empresas se segmentan de acuerdo a su tamaño, que por lo general se determina de acuerdo al nivel de ventas o el capital propio de la empresa, evaluado en un periodo determinado de tiempo. En el caso de Chile, la segmentación por tamaño particiona el universo de contribuyentes en 5 segmentos, que denominaremos C(5):

S1 -> Personas Naturales

S2 -> Grandes Empresas

C(5) S3 -> Medianas Empresas

S4 -> Pequeñas Empresas

S5 -> Microempresas

Se debe considerar que esta segmentación se ha desarrollado definiendo a priori los segmentos que serán construidos, así como los criterios de pertenencia a dichos segmentos, es decir, no es un algoritmo el que define o encuentra estos segmentos, sino que es la Administración Tributaria quien ha decidido tener estos 5 grupos o segmentos de contribuyentes, así como las condiciones o reglas que determinarán la pertenencia a cada uno de ellos.

Desde el punto de vista de negocio, es posible que esta segmentación haya nacido producto del benchmarking entre Administraciones Tributarias, a partir de una tendencia mundial o haya surgido a partir de un estudio que determinó la importancia de definir estrategias diferenciadas por tipo y tamaño de contribuyente. Ahora bien, cualquiera haya sido el origen de esta forma de partición, es importante notar que el objetivo es diseñar estrategias diferenciadas porque se estima que estos contribuyentes tienen necesidades diferentes o enfrentan complejidades diferentes para cumplir o utilizan esquemas diferentes para planificar o minimizar sus cargas tributarias. Ahora bien, las estrategias no tienen por qué ser radicalmente diferentes, pueden ser matices, siempre existirán algunas soluciones estándares y estará también la componente recursos disponibles, como limitante. Por otro lado, un elemento muy importante a tener presente con este tipo de segmentación, guarda relación con las fronteras entre los diferentes segmentos, esto ocurre al exigir o establecer reglas de segmentos en base a variables numéricas continuas, lo que en muchos casos implica que se dé por ejemplo que, los contribuyentes más grandes dentro del segmento de Medianas Empresas podrían tener niveles de venta o capital propio similar al de los contribuyentes más pequeños del segmento de Grandes Empresas, lo que refuerza la mirada sobre los aspectos fundamentales en cuanto estrategias diferenciadas.

Los atributos que permiten definir a cada uno de los segmentos S_k que particionan $C(5)$ se pueden resumir en 7, considerando una simplificación de las definiciones del Servicio de Impuestos Internos, para efectos de ejemplificar:

- i. Atributo 1 sobre tipo de persona que se trata el contribuyente que puede ser obtenida de la base de datos de registro de contribuyentes.
- ii. Atributo 2 sobre tipo de Impuesto a la renta al que queda sujeto el contribuyente que puede ser obtenida del registro de contribuyentes.
- iii. Atributo 3 sobre el monto de ingresos del contribuyente que puede ser obtenida de sus declaraciones de impuesto.
- iv. Atributo 4 sobre monto del capital propio tributario que puede ser obtenida de sus declaraciones de impuesto.
- v. Atributo 5 sobre el monto desembolsado por el contribuyente en compras que puede ser obtenida de sus declaraciones de impuesto.

- vi. Atributo 6 sobre el origen de capitales que puede ser obtenido del registro del contribuyente o sus escrituras de constitución.
- vii. Atributo 7 sobre pertenencia a una nómina de grandes contribuyentes que debe ser determinado analizando otros factores que se verán más adelante.

Recordando la importancia de las metodologías de minería de datos mencionados en apartado anterior, en particular los procesos de selección, limpieza y transformación de datos, se hace referencia a lo que podría ser la salida del proceso de transformación de atributos, previo al modelamiento:

- i. Atributo 1 (A1) – Tipo de Persona
Atributo categórico, binario, toma los valores 0 o 1.
 - i. 0: Persona natural
 - ii. 1: Persona jurídica o sociedad de personas
- ii. Atributo 2 (A2) – Tipo de Impuesto
Atributo binario, toma los valores 0 o 1.
 - i. 0: Impuesto personas físicas (Impuesto de Segunda Categoría de Chile)
 - ii. 1: Impuesto a las sociedades (Impuesto de Primera Categoría de Chile)
- iii. Atributo 3 (A3) – Ingreso
Atributo numérico, que se transforma a categórico para agregar valor a la segmentación que se busca realizar.
 - i. 1: Mayores o iguales a 60.000 UTM
 - ii. 2: Mayores o iguales a 15.000 UTM y menores a 60.000 UTM
 - iii. 3: Mayor o igual a 1.400 UTM y menor a 15.000 UTM
 - iv. 4: Menor a 1.400 UTM
- iv. Atributo 4 (A4) – Capital propio tributario
Atributo numérico, que se transforma a categórico para agregar valor a la segmentación que se busca realizar.
 - i. 1: Mayores o iguales a 300.000 UTM
 - ii. 2: Mayores o iguales a 75.000 UTM y menores a 300.000 UTM
 - iii. 3: Mayor o igual a 1.350 UTM y menor a 75.000 UTM
 - iv. 4: Menor a 1.350 UTM

v. Atributo 5 (A5) – Compras

Atributo numérico, que se transforma a categórico para agregar valor a la segmentación que se busca realizar.

- i. 1: Mayores o iguales a 60.000 UTM
- ii. 2: Mayores o iguales a 15.000 UTM y menores a 60.000 UTM
- iii. 3: Menor a 15.000 UTM

vi. Atributo 6 (A6) – Origen de capitales

Atributo categórico, binario, toma los valores 0 o 1.

- i. 0: Nacionales
- ii. 1: Extranjero total o parcial

vii. Atributo 7 (A7) – Nómina de grandes contribuyentes

Atributo categórico, binario, toma los valores 0 o 1.

- i. 0: No pertenece a nómina de grandes contribuyentes
- ii. 1: Pertenece a nómina de grandes contribuyentes si cumple al menos uno de los siguientes dos criterios:
 - a. Tener ingresos o ventas anuales iguales o superiores a 90.000 unidades tributarias anuales (UTA), en todos y cada uno de los tres últimos años comerciales.
 - b. Aquellas multinacionales que se encuentren obligadas a presentar el “Reporte País por País”⁸, en Chile o en el extranjero.⁹

8 Declaración jurada 1937 “Reporte país por país” establecida por la resolución N° 126 diciembre de 2016, establece que estarán obligados a presentarla:

“A. Entidad matriz o controladora del Grupo de Empresas Multinacional, en adelante también GEM, que tenga residencia en Chile para los efectos tributarios, en la medida que los ingresos del conjunto de entidades que formen parte de dicho grupo, en los 12 meses anteriores al inicio del período tributario correspondiente, sea al menos de 750 millones de euros al momento del cierre de los estados financieros consolidados, según el tipo de cambio observado al 1 de enero de 20151, determinado por el Banco Central de Chile conforme al N°6 del Capítulo I del Compendio de Normas de Cambios Internacionales; o
B. Entidad que integra o pertenece al GEM, que tenga residencia en Chile para los efectos tributarios, y que haya sido designada por la entidad matriz o controladora de dicho grupo, como la única sustituta de éste, para efectos de presentar la Declaración Jurada “Reporte País por País” en su país de residencia tributaria, a nombre de la entidad matriz o controladora.”

9 Resolución Exenta N° 143 de diciembre de 2019.

5.1.1.1 Reglas de pertenencia a segmento S1

De esta forma la primera segmentación, que particiona el universo en dos, personas y empresas, puede ser determinado con 2 de los 7 atributos, A1 y A2. Evaluando los valores posibles de los atributos A1 y A2, es decir, $\{(0,0); (0,1); (1,0); (1,1)\}$ es posible determinar que para todas las combinaciones de los atributos de un contribuyente i , $X_i = (X_{i1}, X_{i2}, X_{i3}, X_{i4}, X_{i5}, X_{i6}, X_{i7}) = (A1, A2, A3, A4, A5, A6, A7)$ donde A1 y A2 sea igual a 0, estaremos sobre el segmento S1 de Personas Naturales.

5.1.1.2 Reglas de pertenencia a segmentos S2, S3, S4 y S5

La clasificación según tamaño queda determinada por los atributos A3, A4, A5, A6 y A7. En este caso, la pertenencia a cada S_k se determina evaluando el cumplimiento de reglas sobre estos atributos que componen el vector de atributos del contribuyente. De esta forma, considerando que se trata de personas jurídicas, el vector de atributos para cada uno de los contribuyentes “ i ”, considera todas las combinaciones posibles donde los A1 y A2 son distintas de (0, 0)

$$\{ \quad X_i = (1, 0, A3, A4, A5, A6, A7)$$

$$X_i = (X_{i1}, X_{i2}, X_{i3}, X_{i4}, X_{i5}, X_{i6}, X_{i7}) \quad X_i = (0, 1, A3, A4, A5, A6, A7)$$

$$X_i = (1, 1, A3, A4, A5, A6, A7)$$

Desde el punto de vista analítico, esta asignación es relativamente muy simple, por cuanto se basa en la evaluación del cumplimiento de ciertas reglas. Por ejemplo, la regla de pertenencia segmento S2 de grandes empresas, puede ser descrita de la siguiente forma:

$$X_i \in S2 \Leftrightarrow (A3 = 1) \vee (A4 = 1) \vee (A5 = 1) \vee (A6 = 1) \vee (A7 = 1)$$

De cumplirse la regla anterior, el contribuyente “ i ” cuyo vector de atributos es X_i , será clasificado en el segmento S1 de Grandes Empresas. En otras palabras, el contribuyente “ i ” será clasificado en Grandes Empresas, si y solo si, al menos uno de los atributos A3, A4, A5, A6 o A7 toma el valor 1.

De manera similar, se pueden identificar los diferentes vectores para cada uno de los contribuyentes que componen el padrón o registro de contribuyente, aplicar las reglas de pertenencia y determinar el segmento en el que quedará asignado cada uno de los contribuyentes. Dadas las posibilidades de valores que pueden tomar cada uno de los diferentes atributos A1 a A7, existen 576 posibles combinaciones de valores para conformar el vector de atributos para personas jurídicas. Valga notar que, por construcción, dada la correlación entre atributos, algunas de estas combinaciones podrían no ser prácticamente posibles, por ejemplo, es posible que, Ingresos esté correlacionado con Capital. Por otro lado, ambos atributos, 3 y 7, referencian montos, al menos en una componente. Sin perjuicio de lo anterior, se muestra un set de ejemplos de lo que sería el resultado de la aplicación de reglas:

Vector							Segmento	
A1	A2	A3	A4	A5	A6	A7		
1	1	1	1	1	0	0	Grandes	S1
1	1	1	1	1	1	0	Grandes	S1
1	1	2	2	2	0	0	Medianas	S2
1	1	3	3	3	0	0	Pequeñas	S3
1	1	4	4	3	0	0	Micro	S4

5.1.2 Segmentación no supervisada en Grandes Contribuyentes

Considerando la información disponible y las técnicas de minería de datos que se han comentado, el análisis de segmentación cobraría mayor valor que el anterior si indagamos dentro de los segmentos detallados. Por ejemplo, indagar dentro del segmento de los grandes contribuyentes. En el caso de Chile existen del orden de 14 mil grandes contribuyentes y dentro de estos, destacan del orden de mil contribuyentes que son clasificados en la nómina de gestión de la Dirección de Grandes Contribuyentes¹⁰, dada la complejidad de sus operaciones y por representar un grupo que, en términos de volumen de operaciones, representan lo que podríamos denominar un outlier dentro del segmento S2.

10 Resolución Exenta SII N° 143 del 31 de Diciembre del 2019.

Se podría definir como universo el conjunto de contribuyentes que componen el segmento S2 e indagar a partir de un algoritmo no supervisado, K-Means por ejemplo, la existencia de segmentos dentro de S2, es decir, subsegmentos, que puedan apoyar y justificar el diseño de estrategias diferenciadas dentro de S2.

Se debe considerar la importancia de formular adecuadamente el objetivo y las preguntas relevantes que se buscan responder, de tal forma que el trabajo se oriente a dichos aspectos de base. De acuerdo a OECD en un reporte especial de análisis a grandes contribuyentes y estrategias asociadas en 8 países desarrollados se detectan en términos de estrategias:

- i. Mejor entendimiento de las obligaciones tributarias
- ii. Marco legal claro y de aplicación uniforme
- iii. Uso de valorización de riesgo para gestionar el cumplimiento
- iv. Fortalecer la cooperación
- v. Balance entre servicios de calidad, innovadores y control efectivo
- vi. Potenciar capacidades internas, procesos y tecnologías

Desde el punto de vista de las características generales de estos contribuyentes:

- i. Concentran una importante proporción de los ingresos tributarios
- ii. Abarcan una gran amplitud de procesos
- iii. Tienen negocio y asuntos tributarios complejos: Múltiples entidades operativas, Diversidad de negocios de interés, Alto volumen de operación, Gran número de empleados, Operaciones internacionales, Operaciones transfronterizas con partes relacionadas, Características únicas de negocio (banca, por ejemplo), Distribución geográfica amplia, Complejidad de materias tributarias y legales, políticas y estrategias para minimizar carga tributaria, Estructuras financieras y de negocio complejas
- iv. Tienen un importante rol intermediario delegado por las Administraciones Tributarias, agentes retenedores e informantes de operaciones de terceros.
- v. Uso de consultores tributarios especializados o tienen áreas tributarias incorporadas al interior de la empresa
- vi. Generalmente se trata de empresas listadas públicamente, corresponden a sectores estratégicos de cada economía, sectores regulados (financiero, banca, seguros, pensiones o pertenecientes a un mismo grupo económico, internacionales o sociedades extranjeras, establecimientos permanentes.
- vii. Por lo general actividades de comunicaciones, tecnologías y energía, recursos naturales, minería, hidrocarburos, gas.

A partir de la información anterior y otro conocimiento de los grandes contribuyentes es posible construir una importante cantidad y variedad de atributos que determinarán el vector de atributos para el proceso de segmentación. A continuación, se enumeran los que podrían ser estos atributos, se puede visualizar que la variedad y complejidad de estos, estará sujeta a la información disponible y al conocimiento y creatividad del analista. Se indican tipos de atributos para efectos de referencia exclusivamente:

- i. Atributos de tipo 1: nivel de ventas e ingresos, nivel de costos y gastos, nivel de inversiones, si pertenece a un grupo empresarial, chileno o extranjero o mixto, cantidad de empresas en las que participa.
- ii. Atributos de tipo 2: cantidad de clientes y proveedores, volumen de documentos emitidos, operaciones con relacionadas, cantidad de documentos versus facturación, cantidad de países con los que opera (donde están ubicados clientes y proveedores)
- iii. Atributos de tipo 3: Sectores en los que participa y cantidad de sectores económicos estratégicos en los que participa.
- iv. Atributos de tipo 4: Impuestos pagados y cargas efectivas de impuesto determinada como impuestos pagados sobre resultados financieros, montos de agregados o deducciones, rebajas y créditos aplicados.
- v. Atributos de tipo 5: Tipo de sociedad anónima abierta, cerrada, de responsabilidad limitada, si es extranjera o establecimiento permanente de sociedad extranjera.
- vi. Atributos de tipo 6: Cantidad de empleados, nivel profesional y de remuneraciones pagado, servicios de outsourcing, presencia regional, cantidad de sucursales.
- vii. Atributos de tipo 7: Existencia de gobierno corporativo o equipos tributarios internos, eficiencia de los sistemas de control interno y nivel de automatización de procesos internos.
- viii. Atributos de tipo 8: Si ha tenido contingencias tributarias, de tipo formal o planificación tributaria agresiva, si ha sido proactivo o reactivo frente a las contingencias.

Como se ha mencionado, las etapas iniciales de cualquier proceso de modelamiento implican necesariamente un entendimiento del problema, del negocio y luego la selección de los atributos, en esta etapa podrían detectarse atributos que no serán posibles de construir, por ejemplo, porque no existe la información o porque la calidad de esta no es adecuada o porque luego de un análisis se estima que no agregan valor al objetivo buscado. Con posterioridad, se despliega la etapa de limpieza que

corresponde al análisis de la información, en cuanto datos disponibles y calidad, detectando y abordando datos perdidos, anómalos y outliers. A continuación, la etapa de transformación se focaliza en la adecuación o construcción de otros atributos a partir de los existentes, donde se aplican técnicas tales como: *normalizar datos*, para dejar atributos numéricos en rangos acotados; *logaritmizar datos*, para suavizar las diferencias numéricas en los datos (ejemplo edad, en decenas, y sueldos en miles o millones); y *discretizar datos*, para transformar atributos numéricos continuos en categóricos, entre otras.

Una vez definido el set de “n” atributos, se conforma el vector de atributos $X_i = (x_{i1}, x_{i2}, \dots, x_{in})$ y se calcula para todos los contribuyentes “i” que pertenecen al universo analizado (S2 – Grandes Empresas) El algoritmo K-Means, permitirá identificar diferentes segmentos en los que es posible particionar este universo.

5.1.3 Cómo opera el algoritmo K-Means

En términos generales el algoritmo K-Means evalúa la distancia entre los diferentes vectores de atributos de los contribuyentes cuya potencial agrupación se evalúa. De esta forma, mientras más cercanos los puntos, más similares serán a efectos de este algoritmo. Para evaluar la distancia el algoritmo utiliza la distancia Euclidiana citada anteriormente:

$$D(X_i, X_j) = \sqrt{\sum_{k=1}^n (x_{ik} - x_{jk})^2}$$

El algoritmo inicia con dos definiciones arbitrarias:

- i. K, es decir, el número de clústeres o segmentos en los que se particionará el universo.
- ii. Las coordenadas de dichos K clústeres, $\mu(k)$, que se denominan centroides, se seleccionan de manera automática, aleatoria o arbitraria. De tal forma que, S_k serán los K segmentos creados por esta partición, $\mu(k)$ es el centroide del segmento S_k .

A continuación, el algoritmo produce una iteración en la cual:

- i. Se calcula la distancia de cada contribuyente a cada centroide y se asigna a cada contribuyente al clúster con el cual la distancia sea menor.

- ii. Se calcula los nuevos centroides de cada uno de los K segmentos, a partir de la media de los diferentes vectores de atributos de los contribuyentes asignados a cada segmento.
- iii. Se vuelve a calcular la distancia de cada contribuyente a cada nuevo centroide y de esta forma asignarlos al clúster con el cual tenga la distancia menor.
- iv. Se continua esta iteración hasta que no existan diferencias significativas entre los centroides.

Lo que está buscando el algoritmo K-Means es minimizar la suma de los errores cuadrados (SSE), es decir, la suma al cuadrado de las desviaciones de cada uno de los vectores al centroide al cual fue asignado:

$$SSE = \sum_{k=1}^K \sum_{Xi \in Sk} (Xi - \mu(k))^2$$

5.2 Análisis de segmentos según riesgo de incumplimiento

En la lógica de la gestión de cumplimiento basada en riesgos, la clasificación por comportamiento tributario es sin duda una de las grandes innovaciones en cuanto segmentación de contribuyente.

Esta partición por comportamiento tributario busca construir segmentos de acuerdo con el riesgo de incumplimiento de las obligaciones tributarias. El riesgo, según las directrices de la norma ISO 31.000 [7] corresponde al efecto de la incertidumbre sobre los objetivos, en este caso asegurar el correcto cumplimiento de las obligaciones tributarias, con frecuencia se expresa en términos de fuentes del riesgo, eventos, consecuencias y probabilidades. Las consecuencias son el resultado de un evento que afecta los objetivos, pueden ser ciertas o inciertas, puede tener efectos positivos o negativos, directos o indirectos. Se pueden expresar de manera cualitativa o cuantitativa. Por otro lado, la probabilidad es la posibilidad que algo suceda, medida o determinada de manera objetiva o subjetiva, cualitativa o cuantitativa y definida en términos generales o matemáticos, es decir, como una frecuencia en un periodo de tiempo determinado.

De esta forma, el riesgo de cumplimiento se clasifica de acuerdo a dos componentes principales: probabilidad y consecuencias. Es común encontrar en la literatura

una matriz de 25 cuadrantes (Matriz de 5x5) para clasificar el nivel de riesgo de un contribuyente “i”, donde cada cuadrante corresponde a un nivel de severidad del riesgo:

$$\text{Nivel de Riesgo}_i = F(\text{Probabilidad}_i, \text{Consecuencia}_i) = F(P_i, C_i)$$

Donde:

Probabilidad = {Casi certeza, Probable, Moderado, Improbable y Muy Probable}

Consecuencias = {Menores, Moderadas, Significativas, Mayores y Extremas}

La distribución de los niveles de riesgos en la matriz 5x5 puede estar determinada a criterio experto o de acuerdo a un análisis matemático. De esta forma, podríamos distribuir los niveles de riesgos en 4 segmentos: S1 Bajo, S2 Moderado, S3 Alto y, S4 Extremo, como se muestra en la tabla siguiente:

Matriz 5x5 - Niveles de Riesgos

	Menores	Moderadas	Significativas	Mayores	Catastróficas
Casi Certeza	Moderado	Alto	Extremo	Extremo	Extremo
Probable	Moderado	Alto	Alto	Extremo	Extremo
Moderado	Bajo	Moderado	Alto	Alto	Alto
Improbable	Bajo	Bajo	Moderado	Moderado	Alto
Muy Improbable	Bajo	Bajo	Bajo	Bajo	Moderado

Dado el contexto anterior, y desde el punto de vista analítico, para efectos de realizar una partición del universo en consideración al riesgo de incumplimiento, debemos necesariamente abordar las variables de Probabilidad y Consecuencias.

5.3 Modelos de Probabilidad

Corresponde a la probabilidad (P_i) que el contribuyente incumpla sus obligaciones tributarias, como una mirada global o general de su nivel de cumplimiento. Supongamos que esta probabilidad puede ser modelada de acuerdo al comportamiento histórico del contribuyente en base a N obligaciones tributarias, de tal forma que:

- i. Se podría determinar la probabilidad P_{ni} (valor continuo entre 0 y 1) de incumplir cada una de las N obligaciones tributarias para cada uno de los contribuyentes “ i ” obligados a ella, a través de alguno de los algoritmos enunciados en el apartado anterior, ejemplo una Regresión Logística, como veremos enseguida.
- ii. La probabilidad P_{ni} de que el contribuyente “ i ” incumpla la obligación “ n ”, se puede clasificar en alguna de las categorías de Probabilidad {Casi certeza, Probable, Moderado, Improbable y Muy Probable}, de acuerdo a un proceso de discretización (transformación de valor continuo a valor categórico):

- a. Casi certeza $\Leftrightarrow P_{ni} \geq 0,9$
- b. Probable $\Leftrightarrow 0,8 \leq P_{ni} < 0,9$
- c. Moderado $\Leftrightarrow 0,2 \leq P_{ni} < 0,8$
- d. Improbable $\Leftrightarrow 0,1 \leq P_{ni} < 0,2$
- e. Muy Probable $\Leftrightarrow 0,1 < P_{ni}$

- iii. Se podría considerar que cada obligación tiene un peso relativo diferente, consistente con la importancia de la obligación para el adecuado resguardo del Sistema Tributario, ω_n , peso de la obligación “ n ”. Ejemplo, se podría considerar que una obligación administrativa, como mantener información actualizada, es de una importancia relativa menor si se compara con la obligación de declarar correctamente las bases imponibles o presentar los formularios de impuestos respectivo.

Se debe cumplir que la sumatoria de todos los pesos sea igual a 1:

$$\sum_{n=1}^N \omega_n = 1$$

- iv. Finalmente se podría determinar la probabilidad de incumplimiento “ P_i ” del contribuyente “ i ” como una combinación de las probabilidades P_{ni} o de las categorías de probabilidades P_{ni} asociadas a cada una de las N

obligaciones que se evalúan. Supongamos que $Cat(P_{ni})$ es la categoría de probabilidad ({Casi certeza, Probable, Moderado, Improbable y Muy Probable} a la que corresponde esa probabilidad P_{ni} . De esta forma, P_i podría estar determinado como:

- a. La moda de las categorías $\Rightarrow Cat(P_i) = Moda (Cat(P_{ni}))$ con $n= 1, \dots, N$
- b. La categoría del promedio de las P_{ni} . Es decir, $Cat(P_i) = Cat \left(\frac{1}{N} \sum_{n=1}^N P_{ni} \right)$
- c. La categoría del promedio ponderado de las P_{ni} . Es decir, $Cat (P_i) = Cat \left(\sum_{n=1}^N \omega_n \times P_{ni} \right)$, donde ω_n es el peso de la obligación “n”.

5.4 Algoritmo de regresión logística

En este apartado revisaremos de qué manera podríamos determinar la probabilidad de un contribuyente “i” respecto de una obligación “n” (P_{ni}) mediante algoritmos de Regresión Logística.

Como punto de partida, consideremos la obligación de presentar correctamente una declaración de impuestos a la renta, por parte de las personas naturales. Recordar la importancia de la primera etapa de la metodología de descubrimiento de patrones en los datos, en cuanto definir adecuadamente el problema a resolver, su alcance y objetivo buscado, en este caso, se buscará determinar si la declaración de impuestos presentada por el contribuyente, más allá de las descuadraturas evidentes que pudiera tener, cumple un patrón que pueda indicar que existe una subdeclaración de ingresos.

Es importante considerar que las personas naturales no siempre están obligadas a presentar las declaraciones de renta, sino que, depende de los tipos y montos de rentas. Por otro lado, usualmente son de presentación anual, lo que en definitiva reduce la cantidad de datos disponibles para análisis, a diferencia del IVA por ejemplo que son 12 presentaciones por año.

Los siguientes tipos de atributos podrían ser relevantes para modelar este fenómeno, considerar que la metodología de análisis de datos requiere el despliegue de procesos de selección, limpieza y transformación de los datos. Por otro lado, los atributos pueden tener variable temporal diversa, anualizada, mensual, promedio, lo que se desarrolla como parte del proceso de transformación y luego de evaluación del poder explicativo de las variables creadas:

- i. Atributos de tipo 1: identifica si el contribuyente desarrolla una actividad remuneradas y de hacerlo, si es en calidad de independiente, dependiente o ambas, jubilado o rentista.

- ii. Atributos de tipo 2: mide la cantidad de clientes o fuentes de renta tales como sueldos, honorarios, pensiones o rentas del capital. Su frecuencia y varianza en el tiempo.
- iii. Atributos de tipo 3: incumplimientos tales como no declaraciones de impuestos, declaraciones presentadas con inconsistencias o descuadraturas, no pago de giros emitidos por la Administración Tributaria, información desactualizada, no concurrencia a procesos de revisión, revisiones con resultado positivo. Si tiene vigente o ha tenido anotaciones graves tales como querellas o anotaciones preventivas de parte de la administración Tributaria.
- iv. Atributos de tipo 4: monto de impuestos determinados, cantidad de franquicias tributarias que utiliza, carga tributaria efectiva, activos y pasivos, montos y tipos.
- v. Atributos de tipo 5: lugar de residencial, actividad económica o giro de negocios que desarrolla, nivel educacional, tipo de profesión, estado civil, número de hijos, edad. Evolución en el tiempo de esto, por ejemplo, si ha sido empleado dependiente y tiene una carrera profesional extensa o si ha sido independiente, dependiente, rentista y ha tenido un desarrollo profesional breve, menos de 5 años, por ejemplo.
- vi. Atributos de tipo 6: participaciones societarias y capital asociado, nacionales o extranjeras, monto de las participaciones, estabilidad de las participaciones, relevancia en las empresas, giros a los que corresponden.
- vii. Atributos de tipo 7: A partir de toda la información anterior, es posible construir los vectores de atributos de todas las personas naturales que pertenecen al sistema tributario que se analiza.

Ahora bien, debemos considerar que, en este caso, al tratarse de un problema de clasificación, es necesaria una variable dependiente, etiquetada, conocida, para que el algoritmo a aplicar pueda buscar, encontrar y aprender un patrón en los datos.

- i. Variable dependiente: Tomará los valores 0 y 1 y estará compuesta por todas aquellas personas que hayan sido seleccionadas para revisión. Se consignará un 1 si la revisión dio resultado positivo y se detectó una subdeclaración de ingresos. Se consignará 0 en caso contrario, es decir, la revisión no arrojó resultado positivo, no fue posible acreditar una subdeclaración. Muchas veces los problemas de clasificación tienen problema al encontrar los casos positivos (1) Suele ocurrir en las Administraciones Tributarias que los universos revisados son muy inferiores a los universos

totales de contribuyentes. Que un contribuyente no haya sido revisado, no implica que esté correcto, y ese es precisamente un supuesto que no es posible tomar, de ahí que el universo de entrenamiento estará centrado exclusivamente en los casos revisados, para estudiar su comportamiento.

La regresión logística, es un tipo de algoritmo supervisado utilizado usualmente en problemas de clasificación binaria, dentro de los cuales podemos encontrar, por ejemplo, clasificar si el contribuyente pertenece a la clase cumplidor (0) o incumplidor (1) de una obligación tributaria particular. Para el caso analizado, si la declaración de impuestos corresponde (1) o no (0) a un patrón de declaración de impuestos a la renta incorrecta. Con lo cual podemos extrapolar que el contribuyente tiene a su vez un perfil de incumplidor de esta obligación, lo que se podría hacer en base a su última declaración o una serie temporal asociada, tema no menor, pero para efectos del ejemplo, no profundizaremos en esta parte.

Como parte de los modelos supervisados, la regresión logística mide la relación entre una variable dependiente o etiqueta y un set de atributos (vector de atributos X) que se utilizan para describir el fenómeno en análisis, devolviendo la probabilidad de la variable dependiente. En definitiva es como una regresión lineal pero cuyos valores de salida están entre 0 y 1. Por otro lado, la traducción de la probabilidad, valor continuo entre 1 y 0 a una categoría 1 o 0, se realiza a través de un proceso de discretización, como el indicado anteriormente, es decir, la definición de un corte a partir del cual se considera que el contribuyente pertenece a un segmento determinado, por ejemplo, si la probabilidad es superior o igual a 0,75, diremos que la declaración de impuestos cumple el patrón de incorrecta y en consecuencia el contribuyente es incumplidor. En nuestro caso particular, la probabilidad se podrá discretizar en las 5 categorías de probabilidad revisadas anteriormente, {Casi certeza, Probable, Moderado, Improbable y Muy Probable}, con sus respectivos rangos de probabilidad asociados.

La regresión logística utiliza la función sigmoide $f(t) = \frac{1}{1+e^{-t}}$ para relacionar la variable dependiente con el vector de atributos o variables independientes.

Supongamos que $P_i(X_i)$ es la probabilidad de incumplir la obligación tributaria del contribuyente "i", cuyo vector de atributos es X_i , para la regresión logística se utilizará:

$$P_i(X_i) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 \times x_{i1} + \beta_2 \times x_{i2} + \dots + \beta_n \times x_{in})}}$$

Con lo cual tendremos que:

$$\text{Logit} (P_i(X_i)) = \text{Ln} \left(\frac{P_i(X_i)}{1-P_i(X_i)} \right) = \beta_0 + \beta_1 \times X_{i1} + \beta_2 \times X_{i2} + \dots + \beta_n \times X_{in}$$

La razón de las probabilidades $P_i(X_i)$ y $(1-P_i(X_i))$ se denomina Odd Ratio, y en definitiva es lo que se modelará con una regresión lineal, para que a continuación, la función logística transforme el valor devuelto por la regresión lineal, en un valor entre 0 y 1.

Como resultado, para cada contribuyente será posible determinar la probabilidad que su declaración de impuestos esté incorrecta, a partir del vector de atributos X_i que tenga. Con lo cual tendremos (P_{ni}) Si realizamos esto para cada obligación n y se combinan de acuerdo a lo indicado precedentemente, podremos obtener la probabilidad de incumplimiento P_i del contribuyente “ i ”.

5.5 Modelos para determinar las consecuencias

La segunda componente del riesgo es la Consecuencia, de acuerdo a las categorías definidas en este ejemplo, los valores posibles son {Menores, Moderadas, Significativas, Mayores y Extremas}. Existen diversas opciones para evaluar las consecuencias asociadas al contribuyente o al riesgo de incumplimiento de un contribuyente, es importante considerar que las consecuencias pueden ser principalmente de recaudación o imagen institucional, cuantitativas o cualitativas, siendo estas últimas más complejas de determinar. Desde el punto de vista de la recaudación o impacto en el erario nacional, se podrían establecer los siguientes métodos:

- i. Tamaño del contribuyente: en el entendido que, mientras mayor sea el tamaño de un contribuyente, mayor será el impacto que puede generar, tanto desde el punto de vista de impuestos como de imagen. Un contribuyente grande genera una mayor exposición mediática, concentran empleo y explican grandes porcentajes de ventas, exportaciones y devoluciones. De esta forma, se determinan las consecuencias de un contribuyente asignando según su tamaño {muy grandes, grandes, medianos, pequeños, micros} a los cuadrantes de consecuencias {Menores, Moderadas, Significativas, Mayores y Extremas}

Una variante a la anterior, correlacionado a tamaño por definición, es realizar la asignación de consecuencias de acuerdo al nivel de ventas y/o compras. Eventualmente se puede analizar la distribución de la variable en cuanto dispersión y realizar la asignación a niveles de consecuencia

{Menores, Moderadas, Significativas, Mayores y Extremas} de acuerdo a análisis de rango intercuartil, por ejemplo, dividir el intervalo en el cual se mueven los ingresos para un determinado periodo en 5 intervalos, de tal forma que el tramo mayor sea el correspondiente a consecuencias extrema y así sucesivamente hasta el tramo menor sea de consecuencias menores.

- ii. Impuestos Pagados: de manera similar a la anterior, es posible determinar el nivel de consecuencias en relación al nivel de impuestos determinado o efectivamente pagado por el contribuyente. En este caso es posible que exista un supuesto fuerte, cual es, asumir que el nivel de impuestos determinado por el contribuyente es el correcto, luego, clasificar su consecuencia en función de este valor podría implicar asignar una consecuencia baja a contribuyentes que circunstancialmente, por ciclo de negocios o incluso por planificación fiscal agresiva, tienen cargas impositivas muy por debajo de lo que deberían.
- iii. Ingresos o impuestos pagados, aplicar un modelo de minería de datos, por ejemplo, un modelo supervisado de regresión para determinar el nivel de ingresos o impuestos que debería pagar un contribuyente. Es una variante del criterio anterior de “Impuestos Pagados”, que busca suavizar el supuesto que los impuestos están correctamente determinados, buscando alternativamente, una estimación de estos. En este caso podríamos utilizar un set de atributos similar al enumerado en los ejemplos de K-means y Regresión Logística anterior, para grandes contribuyentes y personas naturales respectivamente.

El análisis de regresión (simple o multivariable, lineal o no lineal) podría estimar la variable dependiente, “Y”, impuestos determinados o ingresos, como una función del vector de atributos de los contribuyentes (X_i), de tal forma que, en la forma más general, la regresión quedaría determinada por:

$$Y_i = \beta_0 + \vec{\beta} \times \vec{X}_i \mid Y_i \in IR, X_i \in IR^n$$

$$Y_i = \beta_0 + \beta_1 \times X_{i1} + \beta_2 \times X_{i2} + \dots + \beta_n \times X_{in}$$

La regresión busca minimizar la suma del cuadrado de los residuos, entendidos estos como la diferencia entre el valor muestral de la variable dependiente (etiqueta) y la estimación entregada por la regresión, de tal forma que:

$$\sum_{i=1}^m (Y_i - \hat{Y}_i)^2$$

Los modelos de regresión pueden ser evaluados en términos del coeficiente de correlación múltiples, R^2 , que mide el grado de asociación entre la variable dependiente y el vector de atributos, y en consecuencia el ajuste del modelo será mejor, en la medida R^2 sea más cercano a 1.

Se podría enriquecer el análisis al utilizar como medio de contraste la probabilidad de incumplir una obligación tributaria, modelo revisado en el apartado sobre Probabilidad, de tal forma que el ajuste podría ser más fuerte para aquellos cuya probabilidad de subdeclarar impuestos es superior.

Si bien es cierto, matemáticamente el problema puede quedar resuelto, ser robusto y sofisticado, una estimación de esta naturaleza debe necesariamente atender a aspectos de negocios, posiblemente sea necesario realizar estas regresiones en grupos homogéneos de contribuyentes, segmentos particulares determinados, ejemplos, medianos de cierto tipo de actividad con cierto nivel de desarrollo similar. Adicionalmente, dentro del error de estimación existen componentes económicas más difíciles de incorporar al modelo, tales como tasas de desempleo, movimientos en bolsa, precio de commodities, contingencias sociales o sanitarias, que pueden sin duda tener un efecto sobre los impuestos determinados.

- iv. Estimar la consecuencia a partir del resultado del proceso de segmentación revisado para grandes empresas. En este caso a partir de un modelo no supervisado, K-Means, se construyeron diferentes clústeres dentro del universo de grandes empresas. Los contribuyentes que pertenecen a un mismo segmento, tiene un grado de similitud tal que permitiría presumir que su nivel impositivo debiera a su vez ser similar. De esta forma, podría ser posible analizar la distribución de la variable impuestos, en cuanto distribución para todos los que integran un mismo segmento y, de manera similar al caso de la letra anterior, se podrían estimar las consecuencias en consideración a la distribución de los impuestos determinados dentro de cada segmento, reasignando la consecuencia a aquellos que queden en percentiles inferiores.

- v. Estimar el Customer Life Value (CLV), de manera similar al concepto utilizado por las empresas privadas para determinar el valor de los clientes, se podrían estimar las consecuencias de un contribuyente en términos del valor presente de los flujos futuros que representa para el erario nacional. Esa tarea podría requerir una combinación de modelos supervisados, algunos de ellos se han enunciado en diferentes pasajes de este documento. Por una parte, un modelo supervisado de regresión que estime el horizonte de vida que tendrá el contribuyente. Por otro lado, un modelo supervisado que estime los ingresos futuros hasta el último año de vida útil. Interesante la combinación de modelos por cuanto además se propagan errores, tema que debe ser analizado y ponderado.

6. CONCLUSIONES

A lo largo de ese documento, hemos podido visualizar que la gestión del cumplimiento basada en riesgos, establece una serie de desafíos a las Administraciones Tributarias: mayor conocimiento de los contribuyentes para entender su comportamiento e identificar sus necesidades, fortalecimiento de la innovación en el diseño de acciones de tratamiento que permitan mitigar estos riesgos de incumplimiento. Por otro lado, un manejo inteligente de la información en cuanto, exploración, modelamiento y visualización, y el consecuente desarrollo de competencias de los equipos profesionales.

El avance de las tecnologías de información ha abierto importantes posibilidades, permitiendo el desarrollo de herramientas de almacenamiento y explotación eficiente la información, así como ha permitido el amplio despliegue de técnicas analíticas de modelamiento complejo y el fortalecimiento de la toma de decisiones con potentes herramientas de inteligencia de negocios.

La gestión del cumplimiento basada en riesgos se desarrolla en un ecosistema complejo, tanto por la cantidad de contribuyentes que componen el universo de gestión, como por la variedad de características que los identifican, entre otros, tamaños, actividades económicas, regímenes tributarios, franquicias a las que puedes acceder, ubicaciones geográficas, obligaciones tributarias que les afectan, formas jurídicas y procesos administrativos en los que intervienen.

La segmentación de contribuyentes viene a representar una potente herramienta para identificar grupos de contribuyentes similares entre sí, permitiendo reducir dimensiones al problema de gran escala que es la gestión de cumplimiento y de esta forma, orientar acciones de control o asistencia de manera más focalizada.

La segmentación puede ser llevada a cabo mediante diferentes técnicas analíticas de minería de datos, desde modelos supervisados o no supervisados, que permiten agrupar contribuyentes según similitudes, por ejemplo, el tamaño en términos de ingresos o ventas, o clasificarlos según el riesgo de incumplimiento, a partir de predicciones de no presentación de declaraciones de impuestos, no pago de deudas o incorrecto uso de beneficios tributarios.

Independiente de los modelos que se utilicen, siempre será necesaria la aplicación de metodologías de trabajo (CRISP; SEMMA; KDD) que guíen el actuar de los analistas de datos en la necesaria comprensión del problema, la identificación de los datos disponibles y su consecuente limpieza, la identificación de los atributos o variables relevantes, construcción del set de datos con la transformación de información de por medio, el modelamiento analítico y la interpretación y evaluación de resultados, en cuanto calidad de modelos o análisis de costos o errores.

En este ámbito, se enumeran una serie de posibilidades que se abren en cuanto aspectos de negocio, tales como agrupación para despliegue de medidas de asistencia, determinación de nivel ingresos esperados o pago de impuestos esperados, identificación de patrones de falladores de obligaciones tributarias, clasificación de facturas para determinar procedencia de créditos o predicción del futuro cese de actividades, entre otras potencialidades.

Finalmente, dos ejemplos puntuales con mayor detalle de las técnicas asociadas, para clasificar por tamaño y riesgo de incumplimiento, muestran lo complejo pero posible, dado el desarrollo de las tecnologías de información y herramientas analíticas disponibles en la actualidad, del análisis de información para profundizar el conocimiento de los contribuyentes, modelar sus comportamientos y agruparlos según similitudes, lo que sin duda representa un apoyo muy relevante para la toma de decisiones de las Administraciones Tributarias y la gestión de cumplimiento tributario.

7. BIBLIOGRAFÍA

[1] Servicio de Impuestos Internos - Plan estratégico SII
http://www.sii.cl/sobre_el_sii/plan_estrategico_sii.pdf

[2] Servicio de Impuestos Internos - Plan de Gestión de Cumplimiento Tributario 2019
http://www.sii.cl/sobre_el_sii/plan_cumplimiento_tributario.html

[3] Servicio de Impuestos Internos - Estadísticas
http://www.sii.cl/sobre_el_sii/estadisticas_y_estudios_del_sii.html

[4] Servicio de Impuestos Internos - Cuenta pública
http://www.sii.cl/sobre_el_sii/cuenta_publica.html

[5] Forum on Tax Administration: Compliance management of large business task group guidance note experiences and practices of Eight OECD Countries July 2009 Australia, Canadá, Francia, Irlanda, Holanda, Noruega, Reino Unido, Estados Unidos.
<https://www.oecd.org/ctp/administration/43241144.pdf>

[6] Una Práctica Teoría de juegos – Estrategias para cooperar y competir. Marcos Singer. Ediciones UC. 2013.

[7] ISO 31.000 sobre Gestión de Riesgos – Directrices
<https://www.iso.org/obp/ui#iso:std:iso:31000:ed-2:v1:es>

[8] Documentos Diplomado sobre Inteligencia de Negocios - Universidad de Chile. Profesores Richard Weber, Sebastián Maldonado y Luis Aburto

[9] Documento Técnico N° 70 Implantación, mantención y actualización del proceso de gestión de riesgos en el sector público.
<http://biblioteca.digital.gob.cl/handle/123456789/87>