

Universidad de Buenos Aires
Facultad de Ciencias Económicas
Escuela de Negocios y Administración
Pública

CARRERA DE ESPECIALIZACIÓN EN
MÉTODOS CUANTITATIVOS PARA LA GESTIÓN Y
ANÁLISIS DE DATOS EN ORGANIZACIONES

TRABAJO FINAL INTEGRADOR DE ESPECIALIZACIÓN

Análisis del impacto provocado por los regímenes de
recaudación provinciales en las Pymes argentinas

AUTOR: EMILIANO ESTEBAN LOPEZ FILIPPINI

MENTORA: SILVIA VIETRI

SEPTIEMBRE 2023

Resumen

El Big data está siendo explotado por numerosas organizaciones para su mejora en la productividad y el desempeño general. En el campo tributario los fiscos aprovechan su poder para poder realizar fiscalizaciones sistemáticas permanentes e incluso dar autonomía a ciertas decisiones como, por ejemplo, el envío de notificaciones sin intervención humana.

Los problemas que generan los ingresos brutos de las jurisdicciones argentinas son conocidos por todos los profesionales del campo y lejos de mejorar, cada vez obstaculizan más a las empresas, requiriendo así un estudio constante y exhaustivo de las mejores opciones para los contribuyentes.

A través del aprovechamiento de volúmenes de datos de distintas empresas de un estudio contable de Capital Federal explotado con técnicas de análisis multivariado y *Machine Learning*, este trabajo tiene como objetivo analizar el impacto que tienen los saldos a favor a causa de los regímenes de recaudación desarmonizados entre las provincias. En otras palabras, evaluar sus causas o variables fundamentales y a qué tipo de empresas les impacta más desde la perspectiva financiera y económica. A su vez analiza modelos de clasificación para predecir si los saldos de una pyme van a incrementarse o disminuir en el plazo de 6 meses. Producto de esta información se pretende tener una mirada con mayor granularidad sobre el impacto en las pymes y otorgar algunas observaciones a ser consideradas en la planificación fiscal.

Palabras Clave: Pymes, ingresos Brutos, saldos a favor, planificación fiscal

Índice

Introducción.....	3
Objetivos	6
1. La complejidad tributaria en la era del big data.....	7
1.1. La acumulación de saldos a favor en Ingresos Brutos.....	7
1.2. Big Data en las organizaciones.....	10
1.3. Big Data al servicio del fisco	12
1.4. Big data y planificación tributaria.....	13
2. Obstáculos financieros en Pymes.....	15
2.1. Distintos tipos de pymes.....	15
2.2. La problemática de la financiación en las pymes	16
2.3. El aporte fiscal para empeorar la situación.....	18
3. Análisis de Contribuyentes Pymes.....	19
3.1. Fuente y preparación de los datos.....	20
3.2. Análisis exploratorio y descriptivo preliminar.....	22
3.3. Análisis de Conglomerados de Datos.....	26
3.4. Análisis de componentes Principales.....	30
4. Modelo Predictivo para Pyme con saldo a favor.....	34
4.1. Necesidad de control y previsión de saldos a favor.....	34
4.2. Caso de Pyme de comercio con saldo a favor.....	36
4.3. Modelo Predictivo.....	37
5. Conclusiones.....	41
Referencias.....	42
Apéndice.....	45

Introducción

A todo país le resulta beneficioso la creación de nuevas empresas y / o emprendimientos que promuevan un mayor valor agregado para la economía e incluso el incremento en el nivel de empleo. El producto de esta actividad termina confluyendo en las arcas fiscales a través de los tributos. Así, los sistemas tributarios resultan indispensables para poder llevar a cabo las políticas estatales de cada administración en beneficio de sus ciudadanos. Sin embargo, una pretensión fiscal que no guarde razonabilidad con la capacidad del contribuyente resultará en una evidente desmotivación al dinamismo empresarial y posiblemente en un estancamiento económico.

Argentina se caracteriza por tener una presión tributaria elevada con respecto a la mayoría de los países de Latinoamérica. Sumado a esto, la gran discordia que hay entre las normativas de las distintas jurisdicciones provinciales genera una falta de consistencia y problemas graves para el desarrollo de los contribuyentes. El resultado de esto para las empresas y emprendedores pymes es una mayor necesidad de foco en sus estrategias tributarias. Para ello, debe hacer valer de todas sus opciones y así, llevar a cabo decisiones que impliquen la menor carga de impuestos posible y contar con la mayor información disponible para una toma de decisiones racional y dentro de las opciones de la ley.

Las posibilidades tecnológicas actuales revelan cada vez más datos disponibles pasibles de ser transformados en información. En este contexto las organizaciones han ido asumiendo el uso del Big Data y asimilando su potencial: la gran acumulación de datos, y la experiencia, permite obtener patrones de comportamiento de variables, relaciones entre ellas e incluso el modelado de datos para predicciones. En tal sentido, las organizaciones se están embarcando en un cambio cultural y tecnológico para lograr una transformación que les permita dirigirse a decisiones impulsadas por datos. Es sabido que los organismos fiscales ya están haciéndose de las nuevas tecnologías para mejorar su capacidad fiscalizadora, sin embargo, no existe aún un uso generalizado por parte de los contribuyentes o asesores para mejorar la capacidad de toma de decisiones en este dominio. Resulta entonces imprescindible explorar los usos del Big Data en la toma de decisiones para estrategias tributarias.

El foco principal del presente trabajo es el de aportar una nueva dimensión en los problemas que traen específicamente los regímenes de recaudación de ingresos brutos con la intención también de contribuir a una mejora en la planificación fiscal. Claro está en la doctrina que, ingresos brutos es un impuesto regresivo y plurifásico, y producto de ello, uno de los peores impuestos actuales. Aun así, existe un agravamiento resultante de los adelantos que cobra cada jurisdicción en su afán de lograr recaudación sin apelar a la razonabilidad. Estas políticas devienen en acumulación de saldos a favor de difícil recupero (o imposible) que se mantiene en el tiempo en detrimento de las finanzas de las pymes.

Resulta evidente que la cuestión financiera es de extrema vitalidad para las pymes: según el informe de la confederación argentina de la mediana empresa (s.f.) a partir del IMAF (Índice de medición de acceso al financiamiento) el 98,6% de las pymes encuestadas expresaron la necesidad de financiamiento para invertir y el 49% esgrime que la falta de capital inicial sigue siendo uno de los principales obstáculos para iniciar un negocio sumado a que el contexto económico actual genera incertidumbre. Por lo tanto, también es una obviedad que, acumular saldos con el fisco proporciona un pernicioso obstáculo frente a esta necesidad. Además, esta situación se ve innegablemente exacerbada por el alto impacto inflacionario en Argentina. Como resultado de ello, todo saldo de crédito termina impactando negativamente, no solo en las finanzas, sino en el propio resultado económico de la empresa.

Otra cuestión importante a considerar relacionada con la problemática tratada es la gran expansión del comercio electrónico que hubo durante y posterior al COVID – 19. Producto de ello, las plataformas digitales (ejemplo Mercado Libre) fueron designadas como agente de recaudación por distintas jurisdicciones que, sin tener en cuenta el origen ni el destino de la recaudación, comenzaron a hacerse de adelantos por medio de retenciones. En resumen, el mero hecho de estar vendiendo a través de una plataforma digital implica la desatinada obligación a estar inscripto en todas las jurisdicciones del país. Teniendo en cuenta esto, se generan saldos a favor sin ningún sustento de razonabilidad.

A partir de la problemática enunciada resulta relevante para este trabajo explorar con mayor grado de granularidad los efectos que tienen estas políticas nocivas en las

pymes. Para ello se parte de algunas interrogantes tales como: ¿Resulta lo mismo para dos pymes virtualmente iguales vender por plataforma digital o fuera de ella?, ¿A partir de qué momento las empresas empiezan a sufrir más los efectos de estos saldos con respecto a su patrimonio? ¿Qué variables son más significativas a la hora de acumularse saldo?, ¿Vale la pena apertura el comercio a todas las jurisdicciones? ¿Todas las actividades sufren el mismo efecto? ¿Se pueden tomar medidas antes de que esto suceda? Se parte de la hipótesis de que estos regímenes de recaudación no solo que son nocivos desde un aspecto financiero y económico, sino que también generan diferencias entre pymes virtualmente similares lo que vulnera principios básicos en tributación como la equidad y la igualdad: en pocas palabras que los impuestos deben ser similares para empresas o contribuyentes similares.

El trabajo se organiza en cinco apartados. En el primer apartado se analiza las posibilidades del Big data en el ámbito tributario. Para ello, el trabajo describe la problemática actual en el manejo discrecional de los sistemas de recaudación provinciales por parte de los fiscos jurisdiccionales y su falta de armonización. A su vez, se resume el impacto del *big data* en las organizaciones con una breve reseña de cómo los fiscos ya están implementándolo para una recaudación más eficiente. Por último, se analiza el concepto de la planificación tributaria, cuya relevancia es proporcional a las consecuencias de las medidas deficientes en los sistemas impositivos y la posibilidad de valerse del *big data* para mejorar las decisiones empresariales.

En el segundo apartado se abarca la problemática de la financiación en las pymes. Aquí se encontrarán definiciones y características de las pymes argentinas. Luego se expondrá cómo las decisiones tributarias provinciales entorpecen la disposición de fondos de las pymes y por lo tanto una de sus principales fuentes para la supervivencia.

En el tercer apartado se describe el proceso llevado a cabo para armar el *dataset* que fue utilizada en el presente trabajo. El mismo contiene información sobre 62 contribuyentes pymes de distintas características provenientes de un estudio contable de Capital Federal. Dicha información presenta los saldos a favor en ingresos brutos a diciembre 2022 de las empresas, y las características principales de cada contribuyente en función de variables relevantes para el análisis (Ingresos Anuales, Patrimonio Neto, cantidad de jurisdicciones, etc). Por otro lado, se realiza un análisis de componentes

principales para identificar patrones en el impacto de las variables que más pesan a la hora de acumular saldos a favor en las empresas. Por último, se verificarán que agrupamientos se pueden llegar a conformar a través de un análisis de conglomerados para profundizar en el análisis de impacto por grupos de clientes.

En el último apartado, se propondrá un modelo de clasificación para verificar si es factible predecir si los saldos aumentarán o no en las jurisdicciones. Este modelo será entrenado para una de las empresas que conforman el *Dataset* mencionado anteriormente, para la cual se armará un nuevo *Dataset* con los registros de las declaraciones juradas de los últimos 4 años.

Objetivo general

El objetivo general del presente trabajo es analizar las consecuencias de los regímenes de recaudación de ingresos brutos en las finanzas distintos tipos de pymes Objetivos específicos

Objetivos Específicos

Buscar grupos de empresas con características similares y analizar el impacto financiero por saldos a favor en cada grupo.

Evaluar características de las pymes que las conviertan en más vulnerables con el fin de una mayor precisión planificación tributaria.

Explorar qué variables empeoran la situación de la acumulación de saldos

Verificar si un modelo predictivo puede servir para inferir si los saldos a favor, de una empresa en particular, pueden bajar en el tiempo. Inferir si esos saldos subirán o bajarán en seis meses.

El objetivo general del presente trabajo es analizar las consecuencias de los regímenes de recaudación de ingresos brutos en las finanzas distintos tipos de pymes y evaluar si existen pymes que, por sus características, son más vulnerables a sus efectos para ser tenido en cuenta a la hora de la planificación tributaria. En este orden de ideas, también forma parte del objetivo verificar si es posible utilizar un modelo predictivo para una empresa en concreto, que pueda servir para predecir si los saldos van a bajar o subir en el tiempo. Para ello, como objetivos específicos las propuestas se basan en: Buscar conglomerados en los tipos de empresas analizadas, analizar el impacto en cada grupo,



1821 Universidad
de Buenos Aires

explorar que variables son las que empeoran la situación analizada, inferir si los saldos en una empresa de comercio subirán o bajarán en seis meses.

1. La complejidad tributaria en la era del *big data*

El objetivo de este apartado es describir la problemática de muchas empresas respecto de la acumulación de saldos a favor en sus cuentas con los distintos fiscos provinciales y las posibilidades que el *big data* ofrece para mejorar el proceso de planificación tributaria y en consecuencia morigerar los efectos corrosivos en las finanzas provocados por la acumulación de estos saldos.

1.1. La acumulación de saldos a favor en ingresos Brutos

Un factor cada vez más relevante para el desarrollo de las economías es el funcionamiento del sistema tributario de cada país. La seguridad jurídica que se origine de estos sistemas sin dudas tendrá relevancia en las decisiones económicas de los contribuyentes. A propósito de los tributos comenta Catalina Garcia Vizcaíno (2019) que, “cuando van más allá de ciertos límites razonables, pueden reducir la fuerza productiva de los contribuyentes y resultar ruinoso económicamente, porque destruirían las posibilidades de la prosperidad futura” (p.68).

La obtención de ingresos por parte del Estado es una de sus funciones principales en el ejercicio de su actividad financiera. Al ingresar financiamiento a sus arcas, el Estado puede disponer de recursos para hacer frente al gasto público. Este gasto público se origina en las necesidades públicas que evolucionan con el progreso de la civilización tales como educación, salud, transporte. etc. Como menciona Villegas (1987), los recursos públicos se obtienen principalmente de dos formas (p.7): la primera, es con la explotación del propio patrimonio del Estado, por ejemplo, cuando explota, arrienda o vende sus propiedades o con ganancias de actividades industriales y comerciales. La segunda es recurriendo al patrimonio de los particulares en forma coactiva, mediante los tributos.

Dentro de los tributos, el presente trabajo destaca a los impuestos que, por definición son aquellos donde no existe conexión del obligado con actividad estatal alguna que se singularice a su respecto o que lo beneficie. En otras palabras, como

describe Dino Jarach (2004) "Los impuestos no son precios que paguen los usuarios por los servicios que el estado preste individualmente a cada uno de ellos, tampoco son tributos fundados en una contraprestación por servicios" (p.255).

Entre los impuestos provinciales el más importante a nivel recaudatorio es el Impuesto a los ingresos brutos que grava a los ingresos obtenidos por el desarrollo de actividades económicas. La problemática que conlleva este impuesto es la dificultad de establecer a qué jurisdicción corresponde el gravamen siendo que "En los países federales pueden surgir conflictos de doble o múltiple imposición, entre el ente central y los estados o provincias o entre estos últimos entre sí". (Villegas,1987, p.199). Este conflicto se ha tratado de resolver con el Convenio multilateral, que es un acuerdo al que suscribieron todas las provincias argentinas y, en ese momento, la Municipalidad de la Ciudad de Buenos Aires, a efectos de evitar que un mismo contribuyente que desarrolla actividad de manera habitual en más de una provincia, sea objeto de doble o múltiple imposición del impuesto sobre los ingresos brutos por parte de los fiscos de los territorios donde ejerce dichas actividades.

Pese al Convenio multilateral y las distintas normativas que han surgido para aggiornarse a la realidad económica, existen (y se siguen generando) situaciones en donde no hay seguridad jurídica sobre la aplicación de un criterio determinado ya que los propios fallos de la comisión arbitral son contradictorios. Tal como razona Lenardon (2022) "La única conclusión posible de este tipo de medidas bien intencionadas es un régimen cada vez más complejo, cada vez más distorsivo y cada vez más inseguro. Ergo, será muy difícil para los contribuyentes cumplir con este, puesto que, se haga lo que se haga para intentar estar a derecho, siempre existirá una posible contingencia" (párr. 85). Una de las situaciones controvertidas es el comercio electrónico, que ha facilitado la expansión comercial interjurisdiccional y a la vez se ha desarrollado fuertemente en la pandemia. Sin embargo, este crecimiento ha generado problemas para los contribuyentes a causa de la falta de criterios razonables entre las provincias a la hora de atribuir el ingreso por el impuesto de los ingresos brutos. Como señala Alaniz (2020) la normativa "atenta contra el principio de sustento territorial, ya que crea una ficción de que por utilizar el medio de Internet o similar se cumple con el principio de sustento territorial en el domicilio del

adquirente” (párr. 2), de esta forma por el simple hecho de vender en Mercado libre un contribuyente debería darse de alta en todas las provincias.

A propósito de cómo el fisco recauda los impuestos, existen mecanismos para “adelantar” el cobro de estos, como describe Yatzkaier (2017) los mecanismos actuales son:

a) Régimen de retención: es aquel que hace que un sujeto denominado “agente de retención” sustraiga del monto a cobrar del otro sujeto un importe en concepto del impuesto con el objeto de ingresarlo al Fisco a nombre del sujeto retenido.

b) Régimen de percepción: es el sistema que hace que un sujeto denominado “agente de percepción” adicione al monto a pagar por otro sujeto un importe en concepto del impuesto con el objeto de ingresarlo al Fisco a nombre del sujeto percibido.

c) Régimen de recaudación bancaria: es el sistema en que las instituciones bancarias y financieras, ante un depósito efectuado por un cliente de la entidad, sustraen de su cuenta (corriente, de ahorro, etc.) un monto en concepto de anticipo de un impuesto predeterminado por el Fisco, independientemente de su vinculación con dicha obligación fiscal.

d) Regímenes de anticipos y pagos a cuenta: son sistemas que el Fisco instituye a los efectos de que los contribuyentes paguen a cuenta de determinados impuestos sumas de dinero que serán descontadas de la liquidación final del tributo. (párr. 1)

Estos mecanismos de recaudación exacerban el problema para los contribuyentes, debido a que, al no haber seguridad jurídica y criterios homogéneos entre las provincias, se generan retenciones y/o percepciones en jurisdicciones en donde no se está efectuando una operación. La consecuencia de ello son los saldos a favor, en donde el contribuyente

termina siendo acreedor de las jurisdicciones y teniendo resultados financieros negativos y con los niveles de inflación actuales también perjudicando los resultados económicos.

Como menciona Alaniz (2016) “Los regímenes de recaudación del impuesto sobre los ingresos brutos dejaron hace tiempo de ser sistemas de recaudación en la fuente con el mero propósito de anticipar la obligación fiscal”, ahora los regímenes de recaudación buscan asegurarse recaudar el 100% del impuesto sin armonizar con otras jurisdicciones, evitar la evasión (a costa de no discriminar con un sentido razonable la oportunidad de la retención y/o percepción) y dejando para una instancia posterior la discusión y/o reclamo administrativo de los saldos ingresados por demás por parte de los contribuyentes (y en muchas ocasiones resultando prácticamente imposible poder lograr recuperar esos saldos)”(página 2). Sintetizando esta idea, concluye Repele (2022)” En la Argentina no existe un sistema tributario. Un “sistema” es un conjunto de principios entrelazados con coherencia entre sí. Muchas de las normas argentinas no tienen relación ni coordinación con otras, existiendo muchos parches en ellas, y solo tienen en cuenta el fin recaudatorio, siendo el Convenio Multilateral y las normas dictadas por las provincias un claro ejemplo de ello” (párr. 147). Es de manifiesto todos los problemas que acarrea la situación tributaria actual, sin embargo, es importante verificar si este problema afecta a todas las empresas por igual o existen ciertas empresas que sufren más la situación descrita y obstaculiza el desarrollo normal de los negocios.

1.2. *Big Data* en las Organizaciones

El objetivo de este apartado es el de introducir el concepto del *big data* en las organizaciones, de sus posibilidades para una nueva era de toma de decisiones *data driven* aprovechando el nuevo ecosistema de datos.

Históricamente las organizaciones se han servido de los datos para analizar los porqués de ocurrencias del pasado. La información disponible solía provenir meramente de puertas hacia dentro y su valor confluía en reportes, tableros, informes. El aprovechamiento de esta información permitía tomar medidas correctivas para mejorar los resultados (sobre todo operacionales y financieros) de la organización.

El cambio profundo de los seres humanos en su relación con los datos se fue desarrollando principalmente a partir del nuevo milenio. Desde hace ya más de una década han confluído las redes sociales, la movilidad, aplicaciones, la masividad del uso banda ancha y el internet de las cosas (Aguilar, 2013). Este nuevo contexto en donde nos hemos convertido no solo en usuarios de datos, sino también en generadores masivos de datos conforman el escenario en donde el fenómeno del Big data se ha expandido.

El fenómeno del Big Data tiene fundamentalmente tres características principales: el volumen, la velocidad y la variedad de los datos (Kolanovic y Krishnamachari, 2017). El volumen está relacionado con la gran cantidad de datos pasibles de ser colectados desde distintas fuentes. La velocidad hace referencia a la rapidez en que los datos son enviados y recibidos continuamente pudiendo ser estas transacciones en tiempo real o no. La variedad tiene que ver con las diferentes formas en que los datos están disponibles. Estos datos, a su vez, tienen características y tratamientos diferentes. A mencionar, existen datos estructurados (SQL, CSV), semiestructurados (JSON HTML) o no estructurados (blog post). El aprovechamiento de la explosión de los datos coincide con las nuevas tecnologías para la captura, almacenamiento y procesamiento de estos. Este poder de procesamiento revela otra característica esencial del Big Data: la posibilidad explotar la capacidad de búsqueda, obtener información agregada y cruces de información de grandes volúmenes de datos (Boyd y Crawford, 2012).

Muchas organizaciones ya han asumido la relevancia del uso de datos para su proceso de toma de decisiones. En este aspecto, los estudios son contundentes: las empresas que adoptaron una toma de decisiones basadas en datos son un 5% más productivas y un 6% más rentables (McAfee et al., 2012). Esta toma de decisiones impulsadas por datos provoca grandes cambios en las organizaciones. A saber, se sopesa la intuición de los más experimentados contrastándola con los datos, se generan nuevos roles; y se generan nuevos desafíos, tales como, la adopción de nuevas tecnologías, la forma transversal de la organización de tomar decisiones, cambios culturales, etc. (McAfee et al., 2012).

Existen dos grandes áreas en las organizaciones que explotan el Big Data: la inteligencia de negocios y la ciencia de datos; sus roles en el aprovechamiento de los

datos son complementarios (Schmarzo, 2013). Por un lado, la inteligencia de negocios se enfoca en entender los procesos claves del negocio generando tableros, reportes, alertas para dar soporte a esos procesos. En otras palabras, tiene un foco retrospectivo y agrega valor a partir del análisis de lo sucedido. Por el otro lado, la ciencia de datos busca generar modelos predictivos, escenarios posibles, es decir tiene un foco hacia el futuro o las posibilidades del futuro.

La ciencia de datos contiene el concepto de aprendizaje automático o *machine learning*. El proceso de *machine learning* es parte de un campo más amplio de la ciencia de la computación y la estadística. El objetivo del proceso es aprender de la experiencia con ciertas tareas y de forma autónoma. También puede ser vista como un modelo para reconocer patrones en grandes sets de datos. Las técnicas de Machine Learning incluyen los aprendizajes supervisados (regresiones y clasificaciones), aprendizajes no supervisados (análisis de factores) y aprendizaje profundo. A continuación, será descripto cómo es que ya se están utilizando este tipo de herramientas en el dominio tributario.

1.3. *Big data* al servicio del fisco

Si bien no forma parte del tópico del presente trabajo resulta relevante describir como el Big data y la inteligencia artificial ya forman parte del campo y debate tributario. En el caso argentino, la reciente resolución general 5329/2023 (AFIP) puso en vigencia el “Programa de Monitoreo Fiscal”. Este programa (según el texto de su normativa) controla de manera sistemática, instantánea y permanente las declaraciones juradas presentadas por los contribuyentes. En otras palabras, el fisco nacional ya está utilizando inteligencia artificial (a través de *machine learning*) para llevar a cabo sus funciones con un sistema que de manera autónoma envía comunicaciones a los contribuyentes y genera un reclamo en el organismo.

El uso de la inteligencia Artificial por parte de los fiscos trae como consecuencia debates éticos y legales, después de todo su uso implica (en mayor o menor medida) toma de decisiones sin intervención humana (Rubín, 2023). Además, la opacidad que presentan estos algoritmos a la hora de detectar fraudes o ejecutar alguna acción también resulta de una asimetría de poder de la información hacia el contribuyente. Estas controversias ya forman parte del litigio judicial. Tal es el caso de “Syri” donde el Tribunal fallo en contra

del fisco holandés que, recopilaba datos y elaboraba perfiles de riesgo para la detección de fraudes, por considerarlo ilegal. Al respecto el tribunal expresó “...*SyRI es insuficientemente transparente y verificable. Como tal, la legislación SyRI es ilegal, porque viola la ley superior y, en consecuencia, ha sido declarada sin efectos vinculantes.*”

Uno de los problemas de los fiscos en el cumplimiento de sus funciones es la falta de recursos para procesar la gran cantidad de datos que tienen sobre los contribuyentes (producto de cruces de información con terceros y las propias declaraciones informativas y determinativas). A pesar de las controversias, todo indica que el uso la inteligencia artificial tomará cada vez más relevancia al solucionar las limitaciones de recursos y con el poder de una fiscalización permanente. Teniendo en cuenta que el fisco ya lo está utilizando, la consulta que emerge es ¿Qué posibilidades de uso de estas herramientas hay para los contribuyentes o asesores tributarios?

1.4. Big Data en la planificación tributaria.

El objetivo de este apartado es el definir lo que consiste la planificación tributaria y la posibilidad de uso de herramientas de aprovechamiento de grandes volúmenes de datos mejorar la toma de decisiones en el asesoramiento de las organizaciones.

En todo proyecto humano la planificación se torna relevante a la hora de poder aprovechar las oportunidades que el entorno y las circunstancias ofrecen y evitar los riesgos u obstáculos que también surgen de ellos. En el dominio Tributario, la planificación tributaria puede ser definida como las acciones tendientes a reducir la carga tributaria de las empresas y aumentar su beneficio sin incumplir la legislación vigente. Esta tarea implica el conocimiento de las diferentes normativas, la jurisprudencia y el propio conocimiento en profundidad de las actividades realizadas por la organización.

Para llevar a cabo la planificación tributaria existen ciertos objetivos específicos que se persiguen. A modo enunciativo es plausible nombrar: elegir la estructura societaria más conveniente considerando la incidencia impositiva de cada una de ellas, evitar impuestos innecesarios, conocer negocios alternativos, aprovechar beneficios vigentes, elegir la alocución de la empresa, elegir las jurisdicciones. Todo esto basándose en la no

trasgresión de ninguna normativa para lo que es muy importante distinguir con precisión los conceptos de elusión y evasión fiscal.

La elusión fiscal puede ser definida como:

El conjunto de acciones individuales o combinadas, desplegadas por el contribuyente tendiente a encuadrar sus actos, operaciones y actividades por fuera del hecho imponible, o dentro de un régimen fiscal más favorable evidenciado por la carga tributaria más liviana y/o el diferimiento de la obligación de pago de tributo (Rajmilovich 2013, página 1).

Por el contrario, la evasión fiscal opera contraviniendo la regla del tributo al omitir la declaración del hecho imponible, reducir su base y encuadrar actos, operaciones y actividades dentro de un régimen fiscal más favorable (Rajmilovich, 2013). Dicho de otra forma, ambos conceptos buscan reducir la carga tributaria, sin embargo, las diferencias son muy contundentes en las formas y medios para llevar a cabo dicha tarea.

En cuanto a los ingresos brutos, es posible afirmar que existen ciertas cuestiones importantes a considerar a la hora de planificar: ¿en qué jurisdicciones me conviene operar?, ¿Cómo enmarco mi actividad o que negocio alternativo puedo realizar que reduzca la carga tributaria?, ¿Cómo puedo anticiparme a la acumulación de saldos a favor? ¿En dónde puedo pedir solicitudes de atenuación o exclusión?, ¿Qué consecuencias tendría abrir un canal de ventas por plataforma digital?, etc. Todas estas decisiones y acciones encuadradas en la planificación tributaria tienen su fundamento en la libertad que dispone el artículo 19 de nuestra constitución nacional cuando declara que “Las acciones privadas de los hombres que de ningún modo ofendan al orden y a la moral pública, ni perjudiquen a un tercero, están sólo reservadas a Dios, y exentas de la autoridad de los magistrados”. Por lo tanto, existe un campo de planificación y decisión en los ingresos brutos que resultaría en una mejora en la economía de la empresa.

Por otro lado, la falta de una previsión o planificación respecto de los tributos puede traer consecuencias penosas para las empresas. En el caso de los anticipos y/o pagos a cuentas que se ingresan y generan saldos a favor las consecuencias son directamente financieras e indirectamente económicas, debido a que a medida que avanza el tiempo el costo financiero pasa a ser económico (Alaniz, 2022). Si una empresa tiene

1 millón de pesos de saldo a favor en un mes, el efecto será (simplificando el análisis) simplemente financiero, si es que al otro mes ese saldo se reduce a cero o la empresa tiene saldo a pagar. En cambio, si el saldo es mantenido por un año, los efectos de la inflación acarrearán una pérdida significativa para la empresa.

A propósito de esto, La Vista (2023) opina:

Siguiendo esta misma línea, podemos pensar si los saldos a favor que se generan por anticipación de pago de impuestos, ya sea a través de anticipos, retenciones o percepciones, terminan siendo equitativos o no en un contexto inflacionario. En la actualidad no se pueden ajustar por inflación los saldos a favor del contribuyente, de manera que estos pagos a cuenta de impuestos efectivamente abonados están constantemente expuestos a un efecto pulverizador sobre su valor real. (párr. 4)

Esta situación deja a la luz que, aun cuando dos empresas sean similares, aquella que sufra mayor acumulación de saldos tendrá consecuencias económicas más graves otorgando así, una mayor relevancia a las decisiones a tomar en el proceso de planificación tributaria.

Consecuencia de las distorsiones generadas por la arbitrariedad e incoherencia de ciertas normativas provinciales, resulta de interés evaluar grandes volúmenes de datos para verificar patrones que den respuestas a algunas de las preguntas mencionadas más arriba y que determinen un valor a la hora de asesorar a las empresas en el plano tributario.

2. Los obstáculos financieros en las pymes

El presente apartado tiene como objetivo brindar descripciones y características de las pymes en argentina. Por otra parte, trae a la luz un gran problema para las pymes como lo es la financiación y describe cómo, lejos de cooperar a la supervivencia, los fiscos entorpecen la disposición de fondos para las empresas.

2.1. Distintos tipos de pymes

Si bien el término pymes hace referencia a las pequeñas y medianas empresas resulta imprescindible precisar las características de este tipo de organizaciones. Estas precisiones permitirán distinguir los problemas fundamentales que una empresa en sus rasgos pymes atraviesa. Para ello se tomará como referencia la caracterización que hace el ministerio de producción anualmente que categoriza a las pymes según su facturación o empleo:

Figura 1

Categorización Pymes

Categoría	Construcción	Servicios	Comercio	Industria y minería	Agropecuario
Micro	38.830.000	20.190.000	113.610.000	82.730.000	47.300.000
Pequeña	230.400.000	121.730.000	809.300.000	618.160.000	174.230.000
Mediana tramo 1	1.285.490.000	1.007.530.000	3.846.790.000	4.399.660.000	1.025.360.000
Mediana tramo 2	1.928.020.000	1.438.900.000	5.495.450.000	7.046.710.000	1.626.290.000

Categoría	Construcción	Servicios	Comercio	Industria y minería	Agropecuario
Micro	12	7	7	15	5
Pequeña	45	30	35	60	10
Mediana tramo 1	200	165	125	235	50
Mediana tramo 2	590	535	345	655	215

Fuente: Ministerio de Producción

2.2 La problemática de la financiación en las pymes

Respecto de la importancia de la vida de las pymes en la dinámica económica del país cabe mencionar que, según la secretaria de industria y desarrollo productivo para el año 2021 las pymes representaban un 61% del empleo formal, dividiéndose entre las micro (19,1%), las pequeñas (21,2%) y las medianas empresas (22,2%). En el aspecto de la facturación las pymes concentraron el 46% del total de la economía.

En la matriz de pymes argentinas se puede observar una predominancia en el comercio y las reparaciones:

Figura 2

Distribución de pymes por sector



Fuente: Ministerio de Economía

En conocimiento de la importancia del desarrollo y crecimiento de las pymes cabe preguntarse cuáles son los principales problemas que afrontan las pymes. Este foco resulta gravitante frente al cruel dato que arroja la realidad: argentina tiene la tasa de natalidad empresarial más baja y la tasa de mortalidad más alta¹ de América del Sur (Adler, 2023). De aquellas dificultades numerosas dificultades que se afrontan este trabajo versa sobre la financiación. Como fue mencionado anteriormente, las finanzas y en particular la financiación, resultan indispensable para las pymes. No obstante, la mayoría de los emprendedores carece de conocimientos de finanzas (como debilidad interna) y no resulta tan fácil acceder a buenos créditos para inversiones (como obstáculo externo).

Un estudio del observatorio pymes reveló que entre 2007 y 2017, en promedio, el 60% de los recursos aplicados a la inversión fueron financiados con utilidades (principalmente), aportes de socios y demás recursos propios de la empresa. El 30% que cubre la otra mayoría principalmente se debió a descuentos de cheques.

¹ Solo 1 de cada 10 pymes logra superar los 10 años.

A propósito de los problemas financieros que enfrentan las pymes hoy en día Adler (2023) menciona:

Fuera de la falta de educación financiera, de tácticas y técnicas por parte del gobierno para potenciar y crear emprendedores, como así también las políticas insuficientes para apoyar la exportación que tiene Argentina, vemos también que estamos en un país en donde los bancos sólo pueden prestarle a quien puede demostrar que no lo necesita y eso contribuye al desaliento y la desazón de quienes quieren emprender y necesitan un capital inicial a una tasa baja y a un plazo largo. (párr. 7)

2.3 El aporte fiscal para empeorar la situación

Lejos de cooperar en favor de la supervivencia de las pymes, las políticas fiscales sobre cobros anticipados en general, y en lo particular las provinciales, han provisto de un nuevo marco de imprevisibilidad al funcionamiento financiero de las empresas. Resulta habitual en el ejercicio de la actividad, las empresas se encuentren sistemáticamente incrementando saldos a favor con varias de las jurisdicciones, desviando fondos hacia las arcas del estado, en lugar de destinarlo al circuito operativo o de inversión.

A propósito de ello Chicolino (2020) describe:

Conviven en las 24 jurisdicciones más de 250 regímenes de retención y percepción que obligan a las empresas a disponer de recursos no solo para administrarlos eficientemente sino también para solventar los fondos inmovilizados que generan los ingentes saldos a favor, que para los gobiernos provinciales representan préstamos gratuitos que los contribuyentes les depositan mensualmente. (párr. 6)

Cabe mencionar que existen dos grandes mecanismos para afrontar la situación: Los certificados “no retención y/o no percepción” y los de “exclusión”; y la demanda de repetición de los saldos a favor.

Para la primera solución existen grandes impedimentos para la solicitud de dichos certificados. Algunas jurisdicciones no contemplan plazos de resolución (en una situación donde el tiempo es tremendamente relevante), otras como Tucumán otorgan los certificados por 3 meses (a pesar de lo tedioso que resulta cada presentación de solicitud) y en algunos casos ni si quiera existe la posibilidad de poder agotar todo el saldo debido a las limitaciones que la propia legislación prevé. En este último caso, cabe mencionar a Provincia de Buenos Aires donde la variable dicotómica de tener o no el resultado es si “en el período comprendido por los 3 meses vencidos al mes anterior de la solicitud surja como resultado saldo generado a su favor” (Defilpo, 2022) dejando de lado una variable aún más relevante que es, si existe gran saldo acumulado por periodos anteriores no agotados. Por su parte, CABA no prevé otorgar los certificados de exclusión, en el mejor de los casos es factible una solicitud de “atenuación de alícuota”.

Por su parte, la demanda de repetición es una acción legal que se ve desincentivada por su complejidad ya que involucra la contratación de abogados y especialistas tributarios para iniciar acciones, sin contar de que la empresa pasa a sufrir una serie de inspecciones por parte del organismo demandado hasta la resolución de la misma.

3. Análisis de Contribuyentes

El objetivo de este apartado es extraer conclusiones a partir de los análisis en los datos de clientes de un estudio contable de CABA. Se pretende verificar qué factores exacerbaban la acumulación de saldos a favor, verificar formas posibles de agrupación de empresas, comparar las características de las mismas y analizar el impacto en cada grupo.

Para el análisis se utilizarán dos métodos de Análisis Multivariados. A saber, los dos conceptos claves que componen el análisis multivariado son: las unidades experimentales y las variables. Como define Johnson (2000) “Una unidad experimental es cualquier objeto o concepto que se puede medir o evaluar de alguna manera” (pág. 1). Por otro lado, las variables múltiples aparecen siempre que el investigador analice más de un atributo de cada unidad experimental. Es decir que, las variables son esos atributos o características. De esta manera existen técnicas basadas en las relaciones entre las variables, como, por ejemplo: el análisis de regresión, análisis de componentes principales, análisis de correlación canónica, etc.; y también basadas en las unidades experimentales para encontrar relaciones que pudieran existir entre ellas. En este caso, es

posible citar como ejemplo al Análisis discriminante, análisis por agrupación, etc. Teniendo en cuenta esto y los objetivos comentados se optó por utilizar análisis de conglomerados y componentes principales.

3.1. Fuente y preparación de los datos

Para conformar la base de datos utilizada para el presente trabajo se tomaron en consideración 60 contribuyentes pertenecientes a la clientela de un estudio contable de capital federal. Las empresas y/o empresarios de la muestra tomada son contribuyentes cuya residencia radica en CABA o en Buenos Aires y que en su mayoría tiene actividad en más de una jurisdicción. La muestra excluyó aquellas empresas que tienen regímenes especiales de exención en los ingresos brutos o empresas cuya actividad fuera menor a la de un año. En cuanto a las características de las observaciones, corresponden a empresas de diversos tamaño, patrimonio y rubro.

Los datos para cada una de las observaciones fueron extraídos de los papeles de trabajo de liquidación de ingresos brutos del mes de diciembre del año 2022, en donde se recuperó información sobre el saldo a favor en cada jurisdicción, el alícuota promedio devengada por la organización, las ventas anuales que reflejan el CM05 y el coeficiente correspondiente a cada uno de las provincias involucradas. Por otro lado, se registró como variable, el patrimonio neto de la organización. Esta información provino de los balances legalizados de la misma. En los casos de las organizaciones que no se contaba con balance del año examinado (2022) fue decidido registrar el monto del Patrimonio Neto del año anterior (2021) ajustado por inflación. En aquellos casos que fueran contribuyentes que no presentan balances se decidió aplicar distintos criterios según el caso: en aquellos contribuyentes pertenecientes al rubro de comercio, se tomaron las ventas anuales y se proporcionó a ventas por dos meses (esto representaría un stock de dos o tres meses para el individuo). El resto se aplicó la cifra de 500.000 como una cifra estimada de inversión inicial para un negocio en el año 2022.

Es menester resaltar que, para medir el impacto en las empresas se establecieron dos métricas: el saldo a favor acumulado en proporción a las ventas anuales, y el saldo a favor acumulado en proporción al patrimonio del sujeto. De esta manera es factible

establecer una medida homogénea para estimar daño que se genera en cada caso particular.

A continuación, se describen las variables de la base de datos:

- a. Patrimonio: Representa el Activo – el Pasivo del caso observado del año 2022.
- b. Ventas del Año: Ventas el año a valor histórico, extraído del resumen anual del recálculo de coeficiente (formulario CM05).
- c. Plata Formas Digitales: Variable binaria. Si es 1 representa que vende o cobra a través de Plataformas digitales, tal como Mercado Libre, Mercado Pago, Prisma Medios de pagos, etc.
- d. Promedio de Alícuota: Alícuota promedio entre todas las jurisdicciones y para la actividad principal.
- e. Jurisfuera: Variable dicotómica. Representa como verdadero si el caso observado tiene saldos en Tucumán y/o Misiones, que son las provincias que no adhirieron a Sirtac y Sircreb.¹
- f. Rubro: Categorizado según rubro: 1. Comercio, 2. Servicios, 3. Alquileres, 4. Industria
- g. Tamaño.: Categorizado según tamaño: 1. Micro, 2. Pequeña, 3: Mediana (de acuerdo al cuadro emitido por el Ministerio de producción).
- h. Q Jurisdicciones²: Cantidad de jurisdicciones en las que esta inscripto.
- i. Total: Saldo a Favor total
- j. % de Saldo a favor Ventas: Relación entre el saldo a favor y las Ventas de la unidad observada
- k. % de Saldo a favor Patrimonio: Relación entre el saldo a favor y el patrimonio de la unidad observada.

Como comentario final de este subapartado, cabe aclarar que se decidió no mencionar los nombres de las empresas ni del estudio contable para preservar la confidencialidad y el secreto profesional. Sin embargo, los números representan con

² En una primera instancia, se trabajó con cada jurisdicción como variable y con el saldo a favor respectivo en cada una como registro. Sin embargo, resultó más eficiente a la hora de trabajar los métodos multivariados reducir esa dimensión a la sumatoria en la variable total.

precisión las observaciones en cada una. En el próximo subapartado se describirá el análisis de los datos.

3.2 Análisis Exploratorio Preliminar

Para llevar a cabo el análisis exploratorio y los siguientes análisis se utilizó la plataforma *Colab* de Google y en consecuencia el lenguaje de programación Python (versión 3.6).

Las librerías utilizadas fueron las siguientes: Pandas y Numpy, para el procesamiento y manipulación de la base de datos y análisis exploratorio; Sklearn para trabajos de estandarización de las variables, armado de clúster, armado de modelos predictivos; y Matplot y Seaborn, para realizar los gráficos que se proyectan en este trabajo.

A continuación, se muestra información sobre el análisis de la base datos a través de la librería Pandas y su función *describe*:

Tabla 1

Análisis descriptivo de Dataset

index	Patrimonio	Ventas del año	Plataformas digitales?	Promedio de alícuotas	Jurisfuera	Q de jurisdicciones	Rubro	Tamaño
count	63.0	63.0	63.0	63.0	63.0	63.0	63.0	63.0
mean	167393534.7636	730658742.9102	0.3175	0.0396	0.2222	5.8413	1.9683	2.0794
std	374946790.8404	1959531603.9362	0.4692	0.0149	0.4191	7.2872	1.0621	0.4509
min	-1230000.0	900000.0	0.0	0.002	0.0	1.0	1.0	1.0
25%	2672919.8339	26000000.0	0.0	0.03	0.0	2.0	1.0	2.0
50%	26000000.0	115808144.59	0.0	0.0417	0.0	3.0	2.0	2.0
75%	76433000.0	448000000.0	1.0	0.05	0.0	5.5	2.5	2.0
max	1974000000.0	12332926633.5682	1.0	0.0769	1.0	24.0	4.0	3.0

index	Total	% de saldo Ventas	% de saldo Patrimonio
count	63.0	63.0	63.0
mean	5957032.6668	0.0086	0.0605
std	24546254.346	0.0137	0.1039
min	0.0	0.0	-0.0054
25%	9981.3128	0.0002	0.0002
50%	207539.19	0.0038	0.0161
75%	1944437.0521	0.0121	0.0698
max	189481934.4087	0.0751	0.5199

Fuente: elaboración propia

Lo primero que se verifica en la información de salida es una gran dispersión en los datos. En la fila de desvío estándar los valores para el patrimonio son \$374.946.790 y para las ventas anuales \$1.959.531.603. Esto resulta razonable teniendo en cuenta de que la base está conformada por distintas empresas de distintos rubros y distintos tamaños.

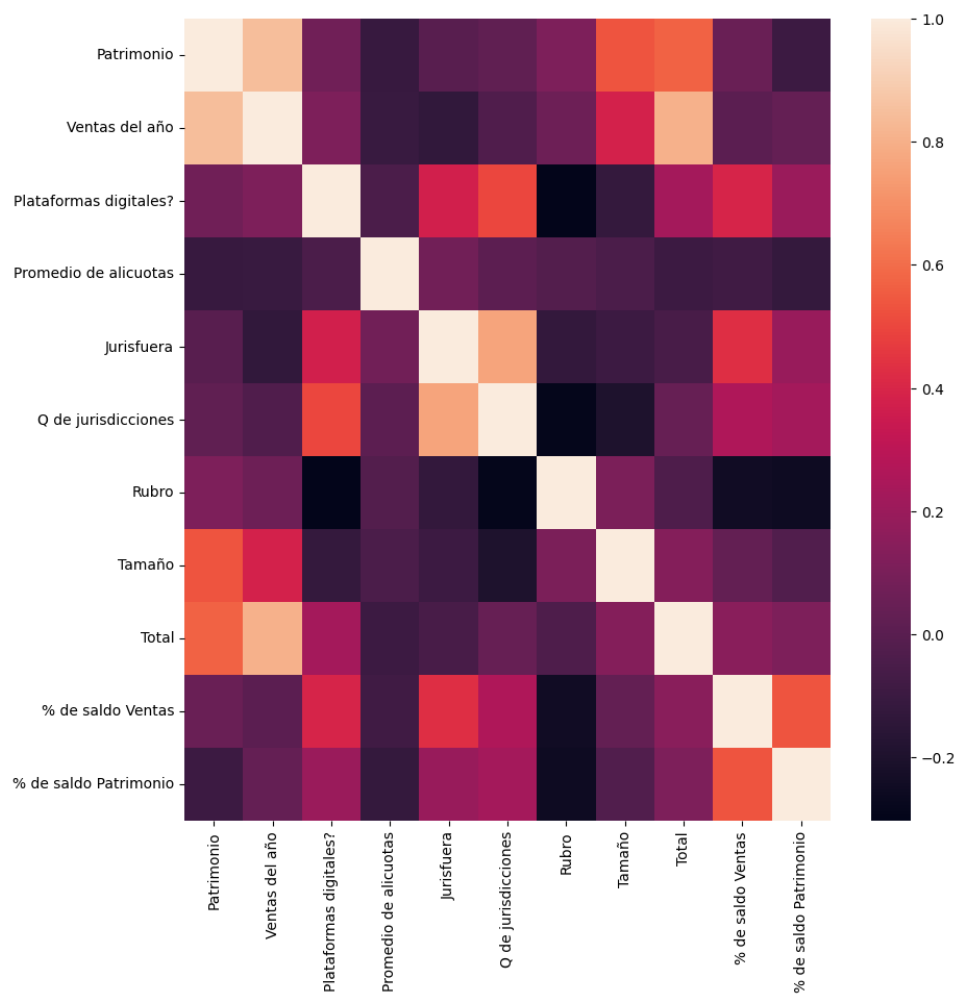
Por otro lado, se verifica que la mediana de la cantidad de jurisdicciones en la que opera cada empresa es 3 y la media es 5,8. Esto se debe principalmente a que el máximo de jurisdicciones a operar es 24 y existen algunas empresas que suben la media al estar inscriptas en todas las jurisdicciones (en general aquellas que realizan comercio electrónico). El rubro predominante resulta Servicios y el tamaño en su mediana y media coinciden en pequeñas empresas.

Un dato que resulta remarcable es que, en la mediana de la proporción de los saldos a favor en función del patrimonio se verifican que más del 1,6% de del monto del patrimonio neto está congelado en un crédito fiscal.

A continuación, se verificó la correlación entre las variables analizadas:

Figura 3

Mapa de correlación de las variables



Fuente: Elaboración propia

A partir del mapa de calor es factible sacar algunas conclusiones preliminares. Es evidente la correlación que hay entre los montos vendidos y el total de saldos a favor acumulados. Cuando se analiza la correlación entre el % de saldos a favor sobre las ventas y el saldo a favor total, esa correlación disminuye considerablemente (0.14). Esto nos indica que vender más, definitivamente propicia a generar más saldos a favor en las jurisdicciones, pero eso no significa que las la acumulación en proporción a las ventas va a ser mayor.

En otro orden de ideas, se verifica que existe un grado de correlación para destacar (0.44) entre la proporción de saldo acumulado en función de las ventas anuales y las empresas que tienen registradas las actividades en Misiones y Tucumán, provincias que no adhieren a Sirtac y Sircreb³. A partir de ello, se puede sospechar que tener abiertas las jurisdicciones Misiones y Tucumán resulta un agravante para la acumulación de saldos a favor de empresas (por lo menos) de Caba y Buenos Aires. Por otra parte, cabe recordar que aquellas empresas que hacen uso de las plataformas digitales tienen adheridas a estas dos provincias, además de que contribuyen a un mayor número de jurisdicciones abiertas, como se ve en el mapa. Esta variable tiene una correlación de 0.2049 con respecto al impacto generado de los saldos acumulados sobre las ventas totales y, si bien no es una fuerte correlación, esa correlación se expresa indirectamente al verificarse que existe una mayor correspondencia entre la cantidad de jurisdicciones adheridas y los saldos acumulados sobre las ventas (0.278).

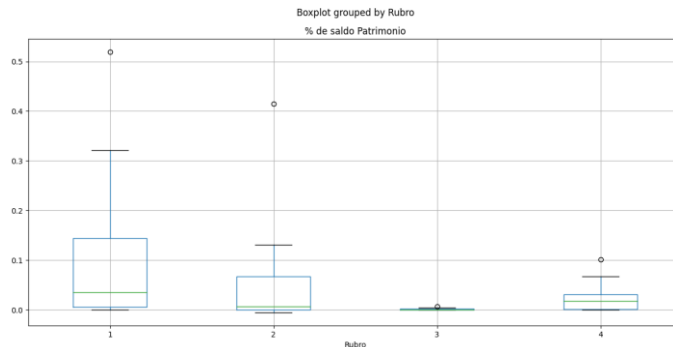
Por último, cabe destacar que existe una correlación entre los dos indicadores que se tomaron para medir el impacto que sufren las pymes a través de las variables % de saldo de ventas y % de saldo de patrimonio. Esta correlación nos da una pauta consistente como métricas de los efectos negativos percibidos por el contribuyente al alterarse el curso de sus finanzas.

A continuación, se analizaron los Outliers en el impacto generado por los saldos a favor por rubro:

³ Estos sistemas gestionan las retenciones de bancos y uso de tarjeta de crédito entre las distintas jurisdicciones aplicando una alícuota unificada de recaudación y posteriormente distribuyendo con un coeficiente entre ellas. Esto pretende evitar que múltiples jurisdicciones apliquen su % de retención a una misma operación.

Figura 4

Boxplot agrupados por Rubro de la variable % de saldo acumulado sobre el patrimonio

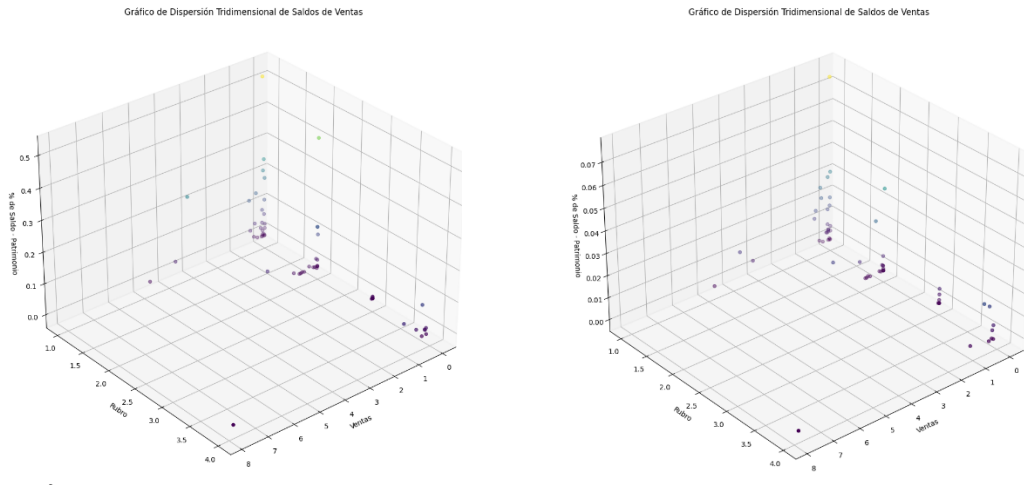


Fuente: Elaboración propia.

Se verifica a partir del gráfico, que los valores de comercio(1) tienden a agruparse en valores más altos que el resto de los rubros y a su vez, tienen una dispersión más grande. En el rubro servicios(2), se encuentra un *Outlier*: al realizar un análisis sobre él se verificó que es un contribuyente proveedor de empresas de construcción, el que recibe los pagos varios meses después del devengamiento de ingresos brutos, y genera por lo tanto saldos a favor por las retenciones sufridas atemporalmente ya habiéndose pagado el impuesto. Aquí se devela otro inconveniente de las retenciones: pese a que el contribuyente ya pago el impuesto correspondiente, posteriormente (en el momento del pago) se le retiene de igual forma. El rubro alquileres (3), sostiene un equilibrio y niveles aceptables de saldo a favor. Es posible argumentar que en general, los alquileres tienen adheridas 1 o 2 jurisdicciones, con lo cual la complejidad y oportunidad para retenciones de diversos orígenes se ve neutralizada. La industria (4), presenta también unos niveles aceptables saldos a favor, pero si tiene un *Outliers*. En un análisis casuístico se detectó que una de esos valores atípicos no resulta de la aplicación sistemática de retenciones sino de errores contables cometidos hace varios años que abrieron el espacio a retenciones que se han transformado en saldos a favor hace mucho tiempo.

Figura 5

Análisis en 3 dimensiones del impacto por rubro y monto de ventas



Fuente: Elaboración propia.

Para continuar explorando los impactos según rubro y monto de ventas, se contrastaron dichas variables con las dos medidas de impacto comentadas anteriormente: El impacto en función de las ventas y en función del patrimonio. En estos gráficos que abarcan las tres dimensiones⁴ se puede observar que, tanto si analizamos los saldos a favor en proporción a las ventas como al patrimonio de la empresa, el impacto adolece más fuerte en empresas del rubro 1, es decir comercio. En otro aporte destacado, se verifica que los efectos más nocivos se dan en empresas que facturaron menos de 150 millones anualmente.

Posteriormente al análisis exploratorio y descriptivo se procedió a los análisis de métodos multivariados, en el próximo sub apartado se tratará dicho análisis.

3.3. Análisis de conglomerados

En el presente sub apartado se describirá el procedimiento y las conclusiones del Análisis de conglomerados.

⁴ Para este grafico se quitó la empresa con mayor facturación (Outlier) para poder enfocar la gráfica en los casos más perjudicados.

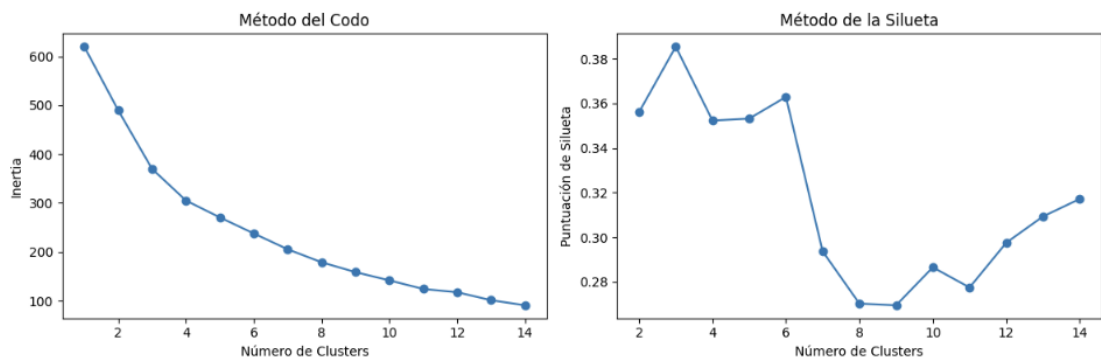
En el caso de los métodos de conglomerados lo que se busca es hacer medición de semejanza o desemejanza entre dos observaciones separadas y luego, nuevamente medir la semejanza o desemejanza entre dos agrupaciones de observaciones. Para ello, se utilizan distintas medidas. Ejemplo de ello es la distancia euclidiana estándar, la distancia métrica estandarizada y la distancia de Mahalanobis.

Para el agrupamiento de los contribuyentes se utilizó el algoritmo de K-medias. El objetivo de este método es el de partir la muestra en k grupos. Cada grupo está representado por el promedio de los puntos que los unen. El representante de cada grupo se denomina centroide. Los centroides comienzan ubicados de forma aleatoria y se asigna a cada observación el centroide más cercano. Luego los centroides se mueven al promedio de todos los datos asignados a él y se vuelven a reasignar las ubicaciones de los puntos de acuerdo con las nuevas posiciones de los centroides.

De acuerdo a lo mencionado lo primero que se realizó fue determinar cuál era el número de k óptimo para determinar los grupos, para ello se realizaron los gráficos de método de codo y de silueta:

Figura 6

Análisis para búsqueda del número de Clúster



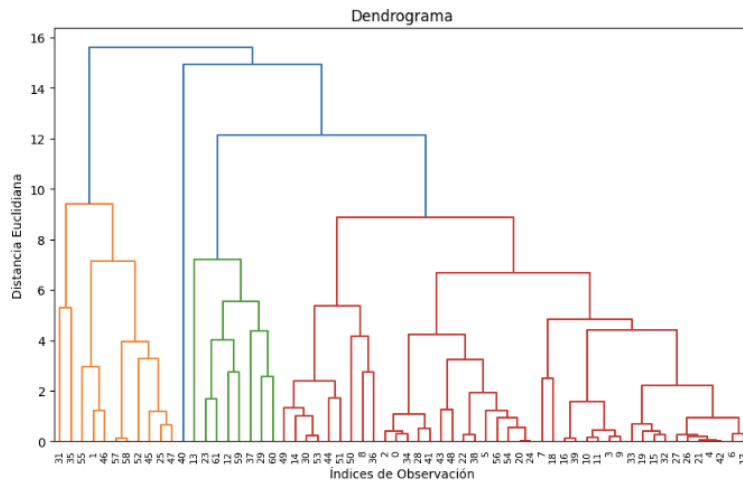
Fuente: Elaboración propia.

El gráfico del codo revela el/los punto/s en el gráfico donde la curva comienza a aplanarse y se asemeja a la forma de un "codo", y ese es el número de clústeres que podría ser considerado óptimo. Por el lado del método de la silueta una puntuación más alta indica que el punto está bien asignado a su clúster y se encuentra lejos de los otros clústeres.

A continuación, también se realizó un dendograma con distancia Euclidiana para ver la posible conformación de grupos.

Figura 7

Dendograma con distancia Euclidiana



Fuente: Elaboración propia.

Como conclusión del uso de estos gráficos se decidió que el número óptimo de clústers era 6. Paso siguiente se procedió a verificar con distancia de Mahalanobis como quedaban conformado esos grupos.

En el grupo número 1 quedo conformada 1 sólo empresa. Esta empresa podría ser considerado un *Outlier* en la muestra ya que es una empresa de negocio de plataforma con operaciones y facturación muy elevados con respecto a los otros valores de la muestra. Tiene operaciones en 7 jurisdicciones y se maneja a través de otras plataformas digitales, en consecuencia, el impacto de los saldos a favor en esta organización mayor que la media del resto de las observaciones.

El grupo numero dos quedo conformado por las siguientes empresas:

Tabla 2

Clúster 2

Índice	Patrimonio	Ventas del año	P.Digitales	Alícuota	Jurisfuera	Q de jurisdicciones
37	855.904.678,80	1.004.823.543,70	0	0,013	1	8
12	71.000.000,00	3.215.037.950,24	0	0,05	0	2
29	719.000.000,00	4.761.655.994,40	0	0,033	0	2
35	836.000.000,00	358.000.000,00	0	0,05	1	16

Índice	Rubro	Tamaño	Total SF	% de saldo Ventas	% de saldo Pat
37	2	3	8.851.706,42	0,0088	0,0103
12	1	3	16.312.151,84	0,0051	0,2297
29	1	3	3.288.686,42	0,0007	0,0046
35	2	3	8.821.800	0,0246	0,011

Fuente: Elaboración propia.

Este grupo de empresas se caracteriza por ser de tamaño medianas, con ventas superiores a los mil millones anuales o patrimonios superiores a los 800 millones de pesos. Tienen un impacto leve al medir los saldos en proporción al patrimonio y comparándolo con la media de la muestra. Se exceptúa de lo comentado al registro 12, cuyo saldo a favor deviene acumulado de una jurisdicción en la cual no opera y por ello no se ha producido una disminución en el tiempo.

En un tercer grupo encontramos a solo dos empresas con patrimonios superiores a los mil millones de pesos. Por un lado, una empresa industrial que no tiene saldos a favor y por el otro una empresa comercial que tiene el 1,1% de sus ventas como saldo a favor y en relación a su patrimonio neto el 3.5%. Dentro de este grupo de empresas patrimonialmente consolidadas existe una diferencia significativa entre el rubro industrial (que suele estar eximido de pagar ingresos brutos o con alícuotas significativamente bajas) y el rubro comercial. Cabe mencionar que la empresa comercial vende también a través de Mercado Libre.

Para un cuarto grupo se revela la siguiente conformación:

Tabla 3

Clúster 3

Índice	Patrimonio	Ventas del año	P.Digitales	Alícuota	Jurisdicción	Q de jurisdicciones
61	213.806.512,50	430.000.000,00	1	0,05	0	3
44	53.600.000,00	385.000.000,00	1	0,05	0	7
48	364.000.000,00	606.661.400,28	0	0,02	0	2
47	291.000.000,00	620.000.000,00	1	0,04	1	2 4
25	63.000.000,00	671.533.000,00	1	0,04	1	2 4
51	332.000.000,00	769.000.000,00	1	0,05	0	6

Índice	Rubro	Tamaño	Total	% de saldo Ventas	% de saldo Pat
61	1	3	11.285.335,88	0,0262	0,0528
44	1	2	8.151.084,69	0,0212	0,1521
48	4	2	11.027.943,87	0,0182	0,0303
47	1	2	10.108.991,64	0,0163	0,0347
25	1	2	8.589.175,56	0,0128	0,1363
51	2	2	50.582,25	0,0001	0,0002

Fuente: Elaboración propia.

Este conglomerado se caracteriza por ser de pymes que facturan entre 400 y 800 millones anuales, casi en su totalidad del rubro comercio y con varias jurisdicciones abiertas. El impacto generado por los saldos a favor respecto del patrimonio es más alto que la media.

Un quinto conglomerado se conforma por una empresa de comercio por plataforma con niveles altos de facturación que no es semejante a ninguna otra empresa de la muestra de acuerdo a los análisis realizados (*Outlier*).

El último grupo está conformado por todo el resto de las empresas de las cuales las 5 que mayor impacto sufren son nuevamente empresas dedicadas al comercio.

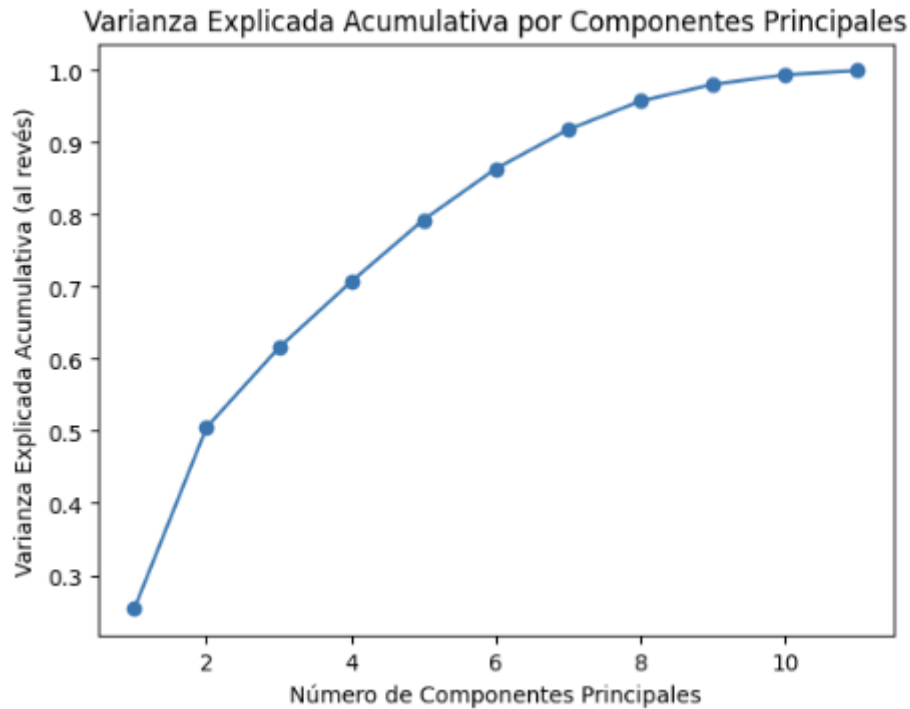
3.4. Análisis de Componentes principales

En el siguiente apartado se describirá el proceso llevado a cabo en el análisis de los componentes principales de la base de datos elaborada. El fin del análisis fue tener una mejor comprensión de la relación de las variables, verificar si existe algún patrón no

evidente y atendiendo a que hay correlación entre las variables, trabajar en la reducción de la dimensionalidad.

Figura 8

Varianza acumulativa de Análisis componentes principales



Fuente: Elaboración Propia

Con la utilización la librería Scikit learn se imprimieron los siguientes resultados:

Figura 9

Varianzas y autovalores para Análisis Componentes principales

```

Varianza explicada:
[25.35345368 25.1764936 11.03835143 9.15958488 8.57960019 7.0875983
5.41844746 3.95671632 2.31311024 1.33054297 0.58610091]

Varianza acumulada:
[0.25353454 0.50529947 0.61568299 0.70727884 0.79307484 0.86395082
0.9181353 0.95770246 0.98083356 0.99413899 1. ]

Autovalores:
[2.83386184 2.81408227 1.23380283 1.02380521 0.9589 7789 0.79221058
0.6056426 0.44225878 0.25854603 0.14872037 0.06551096]

```

Fuente: Elaboración Propia

A partir de la varianza explicada se puede observar que el primer componente no explica significativamente mayor varianza que el segundo, pero entre ambos explican el 50% de la varianza. Por otra parte, al verificar la varianza acumulada se decidió optar por quedarse con 4 componentes ya que explican casi el 70% de la varianza total.

Figura 10

Cargas Factoriales en los primeros 6 componentes:

Loadings por variable en cada componente principal:			
	Componente 1	Componente 2	Componente 3
Patrimonio	0.182111	0.894556	-0.196090
Ventas del año	0.205352	0.936778	-0.059118
Plataformas digitales?	0.724250	-0.027606	-0.121627
Promedio de alicuotas	-0.086644	-0.175701	-0.498458
Jurisfuera	0.712278	-0.282780	-0.423365
Q de jurisdicciones	0.764829	-0.239566	-0.419327
Rubro	-0.463607	0.196774	-0.221592
Tamaño	-0.069280	0.589955	0.032640
Total	0.339046	0.754464	0.027777
% de saldo Ventas	0.702412	-0.056433	0.374401
% de saldo Patrimonio	0.553276	-0.090233	0.618397
	Componente 4	Componente 5	Componente 6
Patrimonio	-0.131974	0.017156	-0.098732
Ventas del año	0.159278	-0.024567	0.084907
Plataformas digitales?	0.268803	0.076741	-0.079757
Promedio de alicuotas	0.015880	-0.808294	0.265566
Jurisfuera	-0.346454	0.132022	0.048507
Q de jurisdicciones	-0.016115	0.194233	-0.090802
Rubro	-0.403878	0.376259	0.595245
Tamaño	-0.606204	-0.222395	-0.403161
Total	0.375401	-0.020241	0.264570
% de saldo Ventas	-0.304893	-0.148197	0.190782
% de saldo Patrimonio	-0.154062	-0.174049	0.254004

Fuente: Elaboración Propia

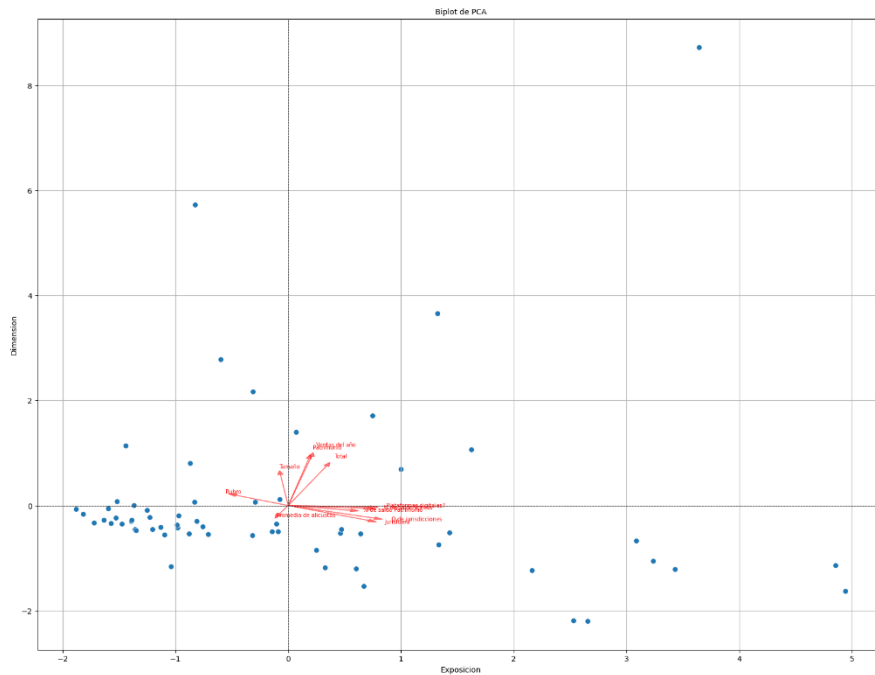
Al analizar con detenimiento cada uno de los componentes se pueden encontrar algunos patrones. El componente 1 podría ser definido como “Exposición en jurisdicciones” ya que tiene una gran carga factorial por parte de ciertas variables relacionadas a un mayor número de actividad en las jurisdicciones (Plataformas digitales, Jurisfuera, Q de jurisdicciones, % de saldo Ventas). En relación al rubro existe una carga factorial negativa de -0.4038, esto se debe a que las empresas más expuestas suelen ser las de comercio y servicios (representadas por los valores 1 y 2, en una escala del 1 al 4). Se verifica en esta relación que las empresas más expuestas suelen ser las de estos rubros. Por su parte, el componente número 2 tiene una gran carga factorial por Patrimonio, ventas del año, tamaño y total de saldo a favor. Este componente puede ser definido como “Dimensión” ya que guarda una relación factorial fuerte con el Tamaño y el monto de operaciones que la empresa realiza. Este componente tiene una carga factorial negativa de -0.24 con respecto a la cantidad de jurisdicciones. Esta relación se explica porque las

empresas más grandes de la muestra revelan una menor apertura de jurisdicciones respecto de otras empresas más pequeñas de la base de datos. El componente número 3 podría considerarse como “Aversión a la Expansión”. Las cantidades de jurisdicciones abiertas, la apertura de Tucumán y Misiones (jurisfuera) contribuyen inversamente en esta dimensión (-0,41). En otras palabras, mientras mayor es la aversión a la expansión de la empresa, menos presencia tiene en el país y en consecuencia menos jurisdicciones abiertas tiene. El valor del patrimonio también tiene una carga factorial negativa en este caso, es decir mientras menos quiera una empresa expandirse menos capitalizará sus ganancias. Siguiendo con el componente 4, se ve un patrón de cargas factoriales relevantes entre rubro y tamaño, es por ello que este componente podría ser asimilado a “Categorización Pyme” ya que justamente las categorías pymes están definidas por el rubro y el tamaño.

En resumen, el análisis de componentes principales revela que la distribución de los datos en el conjunto de empresas estudiadas está fuertemente influenciada por factores como su exposición en plataformas digitales y diferentes jurisdicciones, la dimensión de la empresa, su aversión a la expansión y su categorización dentro del segmento de las pymes. Estos factores, en conjunto, explican aproximadamente el 70% de la varianza total en los datos. Esto sugiere que las características mencionadas desempeñan un papel significativo en la variabilidad observada en nuestro conjunto de datos.

Figura 11

Gráfica de las primeras dos componentes principales, variables y puntuaciones



Fuente: Elaboración propia

A propósito del gráfico Biplot realizado, es posible extraer algunas conclusiones. Aquellos casos observados con coordenada positiva en el eje del componente “Exposición” son empresas que, en sus valores más cercanos al centro, agrupa empresas con condiciones de problemáticas o potenciales problemas respecto de la acumulación de saldo. Algunas, tienen esta potencialidad debido a la cantidad de jurisdicciones abiertas, o por trabajar con comercio electrónico, etc. Otras (las que se encuentran entre los valores 3 y 5), ya tienen problemas propios de acumulación de saldos y a su vez varias jurisdicciones abiertas. Otra observación que resulta de interés resaltar es que, de los casos más alejados al centro (entre los valores 2 y 5 del eje de las abscisas) nos encontramos con empresas “pequeñas” mayoritariamente y 1 solo caso de empresa mediana - grande. Siguiendo en este orden de idea, se verificó que de las 20 empresas más afectadas en relación a su patrimonio 16 tenían patrimonios menores a 100 millones de pesos en el 2022.

En contraposición, el lado izquierdo de las abscisas nos muestra a empresas poco expuestas debido a que: no tienen abiertas numerosas jurisdicciones y/o sus saldos a favor resultan moderados.

4. Modelo Predictivo para Pyme con saldo a favor

El presente párrafo tiene como objetivo analizar si es posible entrenar un modelo predictivo para verificar si los saldos a favor por jurisdicción van a subir o no seis meses hacia adelante en una empresa pyme. Para ello se trabajará con una base de datos de una empresa que se dedica al comercio tanto en la venta al público como a la venta por mercado libre con el objetivo de predecir si los saldos a favor de las posiciones se incrementarán en el plazo de seis meses.

4.1. Necesidad de control y previsión de saldos a favor

Para la gestión impositiva resulta imperativo un seguimiento y control constante. Como se ha mencionado anteriormente, cada una de las jurisdicciones tienen mecanismos a disposición para atenuar los efectos de las retenciones y/o percepciones en exceso. Es por ello, que resulta gravitante mantener un monitoreo sobre los momentos en que sería posible realizar la tramitación de dichas atenuantes y evitar de esta forma un efecto en cascada de mayores saldos a favor imposible de recuperar en el tiempo.

Una de las formas de explotación del Big data es a través del aprendizaje automático. En pocas palabras, los algoritmos de aprendizaje automático se encargan de descubrir patrones en los conjuntos de datos y como subproducto de ello, pueden realizar predicciones con mucha precisión. Para ello resulta relevante la cantidad de observaciones que se tengan registradas, los atributos de las mismas que se relacionan con la variable a predecir, y por su puesto el modelo utilizado y los ajustes de sus Hiperparámetros.

De ello se concluye que resulta una posibilidad para mejorar la planificación tributaria y la previsibilidad financiera de una empresa, la utilización de modelos de aprendizaje que puedan predecir con la mayor exactitud posible el devenir de estos saldos a favor. En el siguiente subapartado se describirá el proceso llevado a cabo para una empresa pyme de Caba con saldos a favor.

4.2 Caso de Pyme de comercio con saldo a favor:

El caso de análisis para este apartado, corresponde a una empresa pyme con apertura en las 24 jurisdicciones dedicada al rubro de venta de electrodomésticos y reparaciones. La empresa reside en CABA y tiene actividad registrada desde hace varios años, sin embargo, la base de datos conformada se compone de las declaraciones juradas desde enero 2020 hasta julio del año 2023 sumando 1027 registros. Cada uno de ellos, se corresponde a la posición de la jurisdicción para cada uno de los periodos con atributos tales como “Coeficiente” (que representa el % atribuible a la jurisdicción), “Anticipo del impuesto Determinado” (el valor del impuesto determinado del periodo), “Saldo a favor anterior” (es el saldo favor remanente del mes anterior) y “a favor fisco” y “a favor contribuyente”. A estos atributos se le agregaron los atributos “inflación” que representa el índice de inflación según Indec de cada mes y “creciouseis” que es una variable dicotómica. En los casos en que en el plazo de seis meses el saldo aumentó el valor de la variable es 1, caso contrario, toma el valor cero.

Figura 12

Dataset Generado en Microsoft Excel para entrenar el modelo de aprendizaje automático

Año	Mes	coeficiente	Jurisdicc	Base Imponible	Saldo a favor anterior	Retenciones	Anticipo del Impuesto Determinado	A favor Fisco	A favor del contribuyente
2020	1	0,3078	901	\$6.218.978,41	\$0,00	\$213.064,03	\$279.854,03	\$66.790,00	\$0,00
2020	2	0,3078	901	\$3.731.007,80	\$0,00	\$49.075,88	\$167.895,36	\$118.819,48	\$0,00
2020	3	0,3078	901	\$1.760.026,14	\$0,00	\$65.379,40	\$79.201,17	\$13.821,77	\$0,00
2020	4	0,3078	901	\$3.018.491,16	\$0,00	\$129.084,88	\$135.832,11	\$6.747,23	\$0,00
2020	5	0,3078	901	\$5.517.279,90	\$0,00	\$295.916,90	\$248.277,59	\$0,00	\$47.639,31
2020	6	0,3078	901	\$5.848.561,09	\$47.639,31	\$326.266,05	\$263.185,25	\$0,00	\$63.080,80
2020	7	0,3078	901	\$6.988.071,07	\$63.080,80	\$299.416,38	\$314.463,20	\$15.046,82	\$0,00
2020	8	0,3078	901	\$4.213.685,09	\$0,00	\$218.538,02	\$189.615,83	\$0,00	\$28.922,19
2020	9	0,3078	901	\$4.358.788,07	\$28.922,19	\$245.742,45	\$196.145,47	\$0,00	\$49.596,98
2020	10	0,3078	901	\$5.486.963,13	\$49.596,98	\$187.476,24	\$246.913,34	\$59.437,10	\$0,00
2020	11	0,3078	901	\$5.769.955,72	\$0,00	\$226.363,12	\$259.648,01	\$33.284,89	\$0,00
2020	12	0,3078	901	\$4.884.427,12	\$0,00	\$192.632,28	\$219.799,22	\$27.166,94	\$0,00
2021	1	0,3078	901	\$4.838.507,21	\$0,00	\$278.221,23	\$241.925,37	\$0,00	\$36.295,86
2021	2	0,3078	901	\$3.558.830,68	\$36.295,86	\$290.728,88	\$177.941,53	\$0,00	\$112.787,35
2021	3	0,3078	901	\$4.379.900,13	\$112.787,35	\$112.787,35	\$218.995,01	\$0,00	\$157.007,57
2021	4	0,3078	901	\$5.246.062,04	\$157.007,57	\$405.527,45	\$262.303,10	\$0,00	\$143.224,35
2021	5	0,3078	901	\$8.586.294,89	\$143.224,35	\$143.224,35	\$429.314,75	\$0,00	\$178.101,00
2021	6	0,3078	901	\$9.750.829,01	\$178.101,00	\$645.220,05	\$487.541,45	\$0,00	\$157.678,60

Fuente: Elaboración propia

4.3 Modelo para predicción de saldos a favor

El *dataset* conformado en Microsoft Excel fue trabajado en Python para poder realizar las predicciones. A partir de la limpieza de algunos datos se definió la variable a predecir que fue “creciéseis”, para saber si la empresa en seis meses incrementaría su saldo a favor en esa jurisdicción. Los atributos que quedaron para entrenar fueron: 'Año', 'Mes', 'Jurisdicción', 'Base Imponible', 'Saldo a favor anterior', 'Retenciones', 'Anticipo del Impuesto Determinado', 'A favor Fisco', 'A favor del contribuyente', 'Inflación'. Todo ello, con el fin de dar input a los modelos predictivos.

Los modelos predictivos se clasifican en dos grandes grupos: modelos de clasificación y modelos de regresión. Un modelo de clasificación es una representación matemática o algorítmica que se utiliza en aprendizaje automático y estadísticas para predecir la categoría o clase a la que pertenece un objeto o instancia basándose en sus características o atributos. Por su parte el modelo de regresión tiene como objetivo predecir una variable numérica continua, su salida es un número real.

En una primera instancia se trabajaron las predicciones con modelos de regresión. Sin embargo, los resultados no resultaron de utilidad debido a los grandes márgenes que arrojaba el error medio cuadrático. En respuesta a ello, se trabajó con distintos modelos de clasificación:

- a) **Arboles de clasificación:** Los árboles de clasificación se basan en una estructura jerárquica en forma de árbol en la que cada nodo representa una

decisión basada en una característica particular de los datos. A medida que se recorre el árbol desde la raíz hasta una hoja, se toman decisiones para clasificar una instancia en una de las clases objetivo. Los árboles de clasificación son fáciles de interpretar y pueden ser útiles para visualizar cómo se toman las decisiones de clasificación. Sin embargo, pueden ser propensos al sobreajuste cuando se construyen árboles muy profundos.

b) **Random Forest:** Es un algoritmo de ensamble que combina múltiples árboles de decisión para mejorar la precisión y reducir el sobreajuste. Cada árbol en el bosque se entrena en una muestra aleatoria de los datos de entrenamiento y con un conjunto aleatorio de características. Luego, las predicciones de todos los árboles se promedian o ponderan para obtener una predicción final. Random Forest es conocido por su capacidad para manejar datos de alta dimensionalidad y su resistencia al sobreajuste.

c) **XGBClassifier:** Crea un conjunto de árboles de decisión de manera secuencial, donde cada árbol corrige los errores del anterior. XGBoost utiliza técnicas avanzadas de optimización y regularización para mejorar el rendimiento del modelo y controlar el sobreajuste.

Para tomar una línea base de referencia (Baseline) para el desempeño del modelo se optó por tomar 61% que es el promedio de aumentos de saldo a favor que existen en el dataset.

A continuación, se exponen los resultados arrojados por las pruebas en los 3 modelos en función de la métrica Accuracy(precisión):

Tabla 3

Resultados de la precisión de los modelos entrenados.

Modelo	Accuracy
Random forest	77,99%
XGBClassifier	80,22%
Árbol de decisión	83,61%

Fuente: Elaboración propia.

El algoritmo que mejor desempeño tuvo fue el árbol de decisión con los siguientes ajustes en sus Hiperparámetros: `criterion='gini'`, `max_depth=35`. Luego, las métricas adicionales revelan un buen desempeño general del modelo:

Figura 12

Métricas varias del modelo Random forest entrenado - Salida Colab.

```
Exactitud: 0.8361581920903954
Precisión: 0.8315789473684211
Sensibilidad: 0.8586956521739131
Valor F1: 0.8449197860962567
Puntuación ROC AUC: 0.835230179028133
Matriz de Confusión:
[[69 16]
 [13 79]]
```

Fuente: Elaboración propia.

5. Conclusiones

A lo largo de los párrafos precedentes se fue desentramando la problemática que resulta para las pequeñas y medianas empresas consistente en lidiar con un sistema inestable e incierto como puede ser el sistema tributario interprovincial para la recaudación de los ingresos brutos. Como característica principal, el presente trabajo se alejó de una perspectiva general del problema (de la cual bastante ya se ha tratado en la doctrina y jurisprudencia), para involucrarse en casos concretos de las situaciones que devienen para las empresas más pequeñas, que como se ha comentado, tienen las máximas posibilidades de desaparecer antes de los 10 años. Adicionalmente se mantuvo el objetivo de generar aportes desde el análisis masivo de datos, proponiendo revelar *insights* para tener en cuenta y el uso de herramientas de aprendizaje automático como oportunidad de mejora para la planificación tributaria.

Uno de los primeros ejes propuestos a explorar en este trabajo fue constatar si los regímenes de recaudación tienen un impacto igual en todas las pymes. A partir de lo expuesto en los análisis de conglomerado y componentes principales, es posible negar dicha afirmación. Se verificó en lo analizado, un impacto superior en las empresas del rubro comercio (*ver figura 3*). En un menor grado de impacto quedaron las empresas de servicios, seguidas mucho después por las empresas Industriales. Cabe mencionar que las empresas dedicadas a rentas de suelo, no tienen este tipo de problemas, por la evidente razón que es la radicación en una sola jurisdicción, y logran así, una ventaja frente a esta problemática.

En otro orden de ideas se verifica una correlación importante (0.44) entre el impacto de los saldos a favor (medidos en función de las ventas anuales) y las empresas que tienen registros de actividad en Misiones y Tucumán. Definitivamente tener abiertas comercio en estas jurisdicciones siendo de jurisdicción ajena trae problemas de mayores retenciones, ya que como fue comentado, no forman parte de los sistemas de recaudación unificados (Sircreb y Sirtac). También se verifica una correspondencia no menor entre la cantidad de jurisdicciones adheridas y el impacto que estas reciben por parte de la acumulación de saldos sobre las ventas (0.27). En el análisis de componentes principales, se verificó en la componente principal a la que se denominó “Exposición”, una gran carga factorial por parte de la cantidad jurisdicciones adheridas, junto con el comercio electrónico y tener apertura en Misiones y Tucumán. Otras componentes principales de

relevancia observadas fueron: la dimensión de la empresa, su aversión a la expansión y su categorización dentro del segmento Pymes. Dentro de las empresas más “expuestas”, es claro notar que las empresas que más sufren el impacto son pymes con ventas menores a 150 millones de pesos anuales. Siguiendo a estas empresas, en pymes con facturaciones superiores (entre 400 y 800 millones anuales) dedicadas al comercio se observa también una de acumulación de saldo e impacto superior a la media (especialmente aquellas que realizan comercio electrónico). Por su parte en las empresas que superan los 800 millones de pesos en sus patrimonios que el análisis de clúster distinguió, se verifica un impacto menor que la media.

En el último apartado se propuso un modelo para entrenar el *dataset* de una empresa que viene teniendo saldos a favor para poder predecir si los saldos en cada una de las jurisdicciones iban a subir o bajar. A partir de ello se pudo observar que el modelo de Árboles de decisión tuvo un buen desempeño y es factible de ser utilizado como herramienta de planificación, previendo los momentos para pedir certificados de atenuación o exclusión, o simplemente para ordenar las finanzas de la empresa.

Resulta de interés en futuros trabajos seguir explorando estas distinciones entre empresas con *datasets*, más grande e involucrando variables que impactan negativamente en las empresas, que fueron dejadas de lado en este trabajo. Un ejemplo de ello es la nominación como agente de recaudación en jurisdicciones en las que la empresa no está radicada, solo por prácticamente tener actividad allí.

Como conclusión final, es posible inferir que los regímenes de recaudación no solo que son malos en las distorsivas e irracionales posiciones en que se fundamentan, sino que además no tratan a todos por igual y aportan un peso adicional a la supervivencia de las pymes llevando al sinsentido (justificado obviamente como ya se observó) de que muchas empresas no quieren vender en ciertas provincias. Tener una empresa de comercio nueva y vender en todas las jurisdicciones definitivamente es una alarma roja que hay que tener en cuenta para la planificación tributaria y financiera. A lo sumo quien emprenda ese camino deberá tener en cuenta que sus márgenes se verán acotados por saldos incobrables o congelados, o en su defecto deberá trasladar este problema al consumidor final elevando sus precios; Por último, el uso de modelos de aprendizaje automático puede ayudar a los planificadores a prever acciones para paliar esta problemática.

Referencias

- Alaniz, J.A. (2022). *Los regímenes de retención y percepción de las 23 provincias y CABA*. Errepar
- Alaniz, J.A. (junio 2020). *Convenio Multilateral. Internet y el Gasto presunto*. Errepar.
<https://eol.errepar.com/sitios/ver/html/20200529181744184.html?k=Internet%20y%20el%20Gasto%20presunto%20alaniz>
- Boyd, d. and Crawford, K. (2012). Critical Questions for Big Data, Information, Communication & Society 15(5): 662-679.
- Confederación argentina de la mediana empresa (s.f). *Relevamiento imaf 2023 sobre medición del acceso al financiamiento de pymes y emprendedores*.
<https://www.redcame.org.ar/novedades/13133/resultado-relevamiento-ima-f-2023-sobre-medicion-del-acceso-al-financiamiento-de-pymes-y-emprendedores>
- Defilpo, N.B (septiembre 2022). *Impuesto sobre los ingresos Brutos. Saldos a favor*. Errepar.
<https://eol.errepar.com/sitios/ver/html/20220810120608212.html?k=saldos%20a%20favor%20ingresos%20brutos>
- Constitución Nacional de Argentina [Const]. Art. 19. 15 de diciembre de 1994 (Argentina).
- Fenochietto, R. (2006). *Economía del Sector Público*. La ley
- Joyanes Aguilar, L. (2013). *Big Data: Análisis de Grandes Volúmenes de Datos en Organizaciones*. México: Alfaomega.
- Johnson Dallas, E. (2000). *Métodos Multivariados Aplicados al Análisis de Datos*. Internacional Thomson Editores.

Kolanovic, M., & Krishnamachari, R. T. (2017). *Big data and AI strategies: Machine learning and alternative data approach to investing*. JP Morgan Global Quantitative & Derivatives Strategy Report.

La Vista, M. (28 de marzo de 2023). *Las inequidades impositivas que provoca la inflación en la Argentina*. <https://www.lanacion.com.ar/economia/las-inequidades-impositivas-que-provoca-la-inflacion-en-la-argentina-nid26032023/>

Lenardon, F. R. (Agosto 2022). *Convenio Multilateral de 1/8/1977: ¿Cuántas comisiones arbitrales existen?*. Errepar. [https://eol.errepar.com/sitios/ver/html/20210701141009414.html?k=Lenardon,%20F.%20R.%20\(2022\).%20Convenio%20Multilateral%20de%201/8/1977](https://eol.errepar.com/sitios/ver/html/20210701141009414.html?k=Lenardon,%20F.%20R.%20(2022).%20Convenio%20Multilateral%20de%201/8/1977)

McAfee, A., Brynjolfsson, E., Davenport, T. H., Patil, D. J., & Barton, D. (2012). *Big data: the management revolution*. Harvard business review, 90(10), 60-68.

Ministerio de economía argentina. (2021). *Estructura y dinamica reciente de mipymes empleadoras*. https://www.argentina.gob.ar/sites/default/files/2021/10/dt_7_-_estructura_y_dinamica_mipyme.pdf

Observatorio Pyme. (s.f.). *Las PyME diversificaron las fuentes de financiamiento durante la última década*. <https://www.observatoriopyme.org.ar/mito/%E2%80%9Cclas-pyme-diversificaron-las-fuentes-de-financiamiento-durante-la-%C3%BAltima-d%C3%A9cada%E2%80%9D>

Rajmilovich, D. (2013). *Planificación fiscal internacional*. La ley

Repele, A. (Marzo 2022). *¿Por qué muchas Pymes no quieren vender en otras provincias ajenas a su sede?*. Errepar. <https://eol.errepar.com/sitios/ver/html/20211101035526706.html?k=Por%20qu>



1821 Universidad
de Buenos Aires

%C3%A9%20muchas%20Pymes%20no%20quieren%20vender%20en%20otras
%20provincias%20ajenas%20a%20su%20sede

Rubin, S. (25 de abril de 2023). AFIP: Inteligencia artificial para el control de evasores.

<https://www.errepar.com/afip-inteligencia-artificial-para-controlar-evasores>

Schmarzo, B. (2013). Big Data: Understanding how data powers big business.

Indianapolis: John Wiley & Sons.

Vizcaíno, C. G. (2019). *Manual de Derecho Tributario*. Abeledo perrot

Villegas, H. B. (1987). *Curso de Finanzas. Derecho financiero y tributario*. Depalma

Yatzkaier, C.E y otros (Junio 2017). *Acerca de los Regímenes de Recaudación e*

Información. Errepar. <https://eol.errepar.com/sitios/ver/html/20170512114137341>

.html?k=Acerca%20de%20los%20Reg%3ADmenes%20de%20Recaudaci%

C3%B3n%20e%20Informaci%C3%B3n

Apéndice

Código utilizado para las predicciones realizadas en *colab* con los modelos entrenados

```
import sklearn as sk
from sklearn import model_selection
from sklearn import tree
from sklearn import metrics
from sklearn.linear_model import LinearRegression
from sklearn.neighbors import KNeighborsRegressor
from sklearn.model_selection import GridSearchCV
import pandas as pd
import numpy as np
from sklearn.model_selection import cross_val_score
from sklearn import ensemble
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.model_selection import cross_val_predict
from sklearn.ensemble import RandomForestRegressor
from sklearn.metrics import mean_squared_error
from sklearn.metrics import mean_absolute_error
from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import accuracy_score
from xgboost import XGBClassifier
from sklearn.metrics import accuracy_score
from sklearn.model_selection import GridSearchCV
from sklearn.metrics import recall_score
from sklearn.metrics import precision_score
from sklearn.metrics import accuracy_score, precision_score, recall_score, f1_score,
    roc_auc_score, confusion_matrix

#APERTURA DE DATASET

df = pd.read_excel("Servi2.xlsx")
df = df[df.columns.drop(["Unnamed: 0", "Unnamed: 3", "Unnamed: 7", "Unnamed:
    8", "Codigo"])]
df['coeficiente'] = df['coeficiente'].fillna(0)
df = df.dropna(subset=['creciouseis'])
X =
    df[df.columns.drop(["creciotres", "creciodos", "creciouseis", "VP6", "VP2", "VP3", "
    coeficiente"])]
y = df["creciouseis"]

##SPLIT TEST TRAIN

X_train, X_test, y_train, y_test = sk.model_selection.train_test_split(X, y, test_size=0.2,
    random_state=42)
```

```
#PRUEBAS DE CLASIFICACION
#Entrenamiento de Arbol de decision
[]
clf = sk.tree.DecisionTreeClassifier(criterion='gini', max_depth=35, random_state=42)
clf.fit(X_train, y_train)

[]
y_pred = clf.predict(X_test)

[]
sk.metrics.accuracy_score(y_train, clf.predict(X_train) )
1.0
[]
sk.metrics.accuracy_score(y_test, y_pred)
0.8361581920903954
[]
sk.metrics.confusion_matrix(y_test, y_pred)
array([[69, 16],
       [13, 79]])
[]
# Evaluamos usando validación cruzada en 5 partes
scores = sk.model_selection.cross_val_score(clf, X, y, cv=200, scoring='accuracy',
                                             n_jobs=-1)
print(scores.mean())
0.76125

[]
param_grid = {
    'n_estimators': [50, 100, 200,250],
    'max_depth': [None, 10, 20],
    'min_samples_split': [2, 5, 10],
    'min_samples_leaf': [1, 2, 4]}
[]
grid_search = GridSearchCV(estimator=clf, param_grid=param_grid, cv=5, n_jobs=-1,
                           verbose=2)

# Realizar la búsqueda de hiperparámetros utilizando GridSearchCV
grid_search.fit(X, y)

# Mostrar los mejores parámetros y el mejor resultado
print("Mejores parámetros:", grid_search.best_params_)
print("Mejor puntuación:", grid_search.best_score_)
Fitting 5 folds for each of 108 candidates, totalling 540 fits
Mejores parámetros: {'max_depth': 10, 'min_samples_leaf': 4, 'min_samples_split': 2,
                    'n_estimators': 50}

# Entrenamiento Random Forest
clf = RandomForestClassifier(n_estimators=150, random_state=42,max_depth=30)
```



```
clf.fit(X_train, y_train)

# Realizar predicciones en el conjunto de prueba
y_pred = clf.predict(X_test)

# Calcular la precisión
accuracy = accuracy_score(y_test, y_pred)

print("Precisión:", accuracy)

scores = cross_val_score(clf, X, y, cv=200)

[]
# Entrenar el clasificador XGBCLASSIFIER

clf = XGBClassifier(n_estimators=300, random_state=42)

clf.fit(X_train, y_train)

# Realizar predicciones en el conjunto de prueba
y_pred = clf.predict(X_test)

# Calcular la precisión
accuracy = accuracy_score(y_test, y_pred)
print("Precisión:", accuracy)
Precisión: 0.8022598870056498
[]
accuracy = accuracy_score(y_test, y_pred)
print("Exactitud:", accuracy)

# Precisión (Precision)
precision = precision_score(y_test, y_pred)
print("Precisión:", precision)

# Sensibilidad (Recall)
recall = recall_score(y_test, y_pred)
print("Sensibilidad:", recall)

Exactitud: 0.8361581920903954
Precisión: 0.8315789473684211
Sensibilidad: 0.8586956521739131
Valor F1: 0.8449197860962567
Puntuación ROC AUC: 0.835230179028133
Matriz de Confusión:
[[69 16]
 [13 79]]
```

Reporte Trabajo Final Integrador de Especialización

“Análisis del impacto provocado por los regímenes de recaudación provinciales en las Pymes argentinas”

Autor: Emiliano Esteban Lopez Filippini

Como mentora del presente Trabajo Final Integrador de Especialización del alumno Emiliano López Filippini, transmito mi opinión sobre la investigación realizada.

Con respecto al tema de análisis, considero que aborda una problemática relacionada con la gestión de datos en organizaciones; en este caso organizaciones relacionadas con el sector tributario de Argentina.

El alumno analiza una base de datos de 62 contribuyentes pymes provenientes de un estudio contable de Capital Federal. En la misma se cuenta con información relativa a los saldos a favor en ingresos brutos a diciembre 2022 de las empresas, y las características principales de cada contribuyente en función de variables relevantes para el análisis, como Ingresos Anuales, Patrimonio Neto, cantidad de jurisdicciones y otras.

El objetivo general es analizar las consecuencias de los regímenes de recaudación de ingresos brutos en las finanzas distintos tipos de pymes, con la intención de contribuir a una mejora en la planificación fiscal.

En la investigación se expone la problemática del gravamen de ingresos brutos para empresas que operan en distintos ámbitos, lo que genera conflictos al respecto de la múltiple imposición y se analiza cómo se ven afectadas las pymes en la disposición de sus fondos, siendo esta una de las causas que pone en juego su supervivencia en el mercado.

El trabajo se enmarca en la gestión de datos en contextos de Big data con el aprovechamiento de grandes volúmenes de datos tendientes a mejorar la toma de decisiones en el asesoramiento de las organizaciones.

Se articulan contenidos de las asignaturas Métodos de Análisis Multivariado (estudio descriptivo de la base de datos y aplicación de métodos multivariantes como el Método de Conglomerados y Método de Componentes Principales) y Modelos Predictivos (modelo de Árboles de decisión)

La coherencia del enfoque planteado, el uso de algoritmos de Minería de Datos, la pertinencia de las referencias bibliográficas y el correcto análisis de los resultados obtenidos, permiten señalar a este trabajo como un aporte relevante en el área tributaria, evidenciando la posibilidad de emprender futuras líneas de investigación.



Dra. Silvia Vietri