



**Facultad de
Ciencias Sociales y Humanísticas**

TESIS

“Estrategias de selección de predictores en contextos de alta dimensionalidad y modelos de aprendizaje automático para el estudio de la adopción de TIC en empresas ecuatorianas”

**Previa la obtención del Título de:
MAGISTER EN CIENCIAS ECONÓMICAS**

**Presentado por:
KATHERINE ESTEFANIE PARRALES GUERRERO**

Guayaquil – Ecuador

2025

AGRADECIMIENTO

A Jesús, mi rey y salvador, por regalarme Su amor, Su bondad y Su paz en cada temporada. La gloria y el honor sea siempre para Él.

A mi familia, por acompañarme durante este proceso. Gracias por no soltar mi mano y animarme con cada palabra de aliento en los momentos más difíciles. Gracias por creer en mí aun cuando la duda invadía mi ser.

A mis mece-amigos por regalarme su amistad y ser una fuente de apoyo y aprendizaje durante este camino.

A mi familia de Más Vida por estar conmigo desde el principio de este reto. Gracias por acompañarme con sus oraciones, con su amistad y con todos sus detalles aun en medio de la distancia.

A mis profesores por ser generosos con sus conocimientos. Gracias por el tiempo dedicado para que pueda aprender y aclarar mis dudas.

A mis colegas de I2MAKER por su apoyo constante. Gracias por todo el aprendizaje recibido.

Al coordinador de esta maestría por darme la oportunidad de ser parte de la cohorte V. Gracias por el acompañamiento y las palabras de ánimo.

A mi tutor de tesis por su valioso respaldo durante todo este tiempo. Gracias por su paciencia y su guía en el proceso de elaboración de este trabajo.

A cada persona que me ha acompañado en este largo trayecto incluso a la distancia. Gracias por regalarme su tiempo, su cariño y por permitirme aprender de cada uno de ustedes. Gracias por inspirarme y marcar una diferencia en mí.

KATHERINE PARRALES

DEDICATORIA

Para Dios y para cada persona que; a través de sus detalles, oraciones, abrazos, palabras de aliento, paciencia, consejos y sabiduría; han sido una curita al corazón en toda circunstancia.

KATHERINE PARRALES

DECLARACIÓN EXPRESA

Yo Katherine Estefanie Parrales Guerrero acuerdo y reconozco que: La titularidad de los derechos patrimoniales de autor (derechos de autor) del proyecto de graduación corresponderá al autor o autores, sin perjuicio de lo cual la ESPOL recibe en este acto una licencia gratuita de plazo indefinido para el uso no comercial y comercial de la obra con facultad de sublicenciar, incluyendo la autorización para su divulgación, así como para la creación y uso de obras derivadas. En el caso de usos comerciales se respetará el porcentaje de participación en beneficios que corresponda a favor del autor o autores. El o los estudiantes deberán procurar en cualquier caso de cesión de sus derechos patrimoniales incluir una cláusula en la cesión que proteja la vigencia de la licencia aquí concedida a la ESPOL.

La titularidad total y exclusiva sobre los derechos patrimoniales de patente de invención, modelo de utilidad, diseño industrial, secreto industrial, secreto empresarial, derechos patrimoniales de autor sobre software o información no divulgada que corresponda o pueda corresponder respecto de cualquier investigación, desarrollo tecnológico o invención realizada por mí/nosotros durante el desarrollo del proyecto de graduación, pertenecerán de forma total, exclusiva e indivisible a la ESPOL, sin perjuicio del porcentaje que me/nos corresponda de los beneficios económicos que la ESPOL reciba por la explotación de mi/nuestra innovación, de ser el caso.

En los casos donde la Oficina de Transferencia de Resultados de Investigación (OTRI) de la ESPOL comunique al/los autor/es que existe una innovación potencialmente patentable sobre los resultados del proyecto de graduación, no se realizará publicación o divulgación alguna, sin la autorización expresa y previa de la ESPOL.

Guayaquil, 24 de septiembre del 2025.

Katherine Parrales

Guerrero

COMITÉ DE EVALUACIÓN

Ph.D. Milton Ismael Paredes Aguirre
Tutor de la Tesis

Ph.D. Holger Barriga Medina
Evaluador 1

Ph.D. Fabricio Zanzzi Díaz
Evaluador 2

ÍNDICE GENERAL

AGRADECIMIENTO	1
DEDICATORIA	2
COMITÉ DE EVALUACIÓN	4
DECLARACIÓN EXPRESA	3
ÍNDICE GENERAL	4
RESUMEN	6
ÍNDICE DE ILUSTRACIONES	7
ÍNDICE DE TABLAS	8
ABREVIATURAS	10
1. Introducción.....	11
2. Revisión Literaria	12
2.1. Relevancia de las TIC en las empresas	13
2.2. Uso de técnicas de aprendizaje automático en la investigación sobre las TIC	14
2.3. Brecha en la literatura	15
3. Metodología.....	16
3.1. Datos	16
3.1.1. Construcción de la base	16
3.1.2. Estrategia de ponderación y representatividad	16
3.1.3. Variable de interés	17
3.2. Selección de predictores	17
3.2.1. Regularización	17
3.2.2. Selección Manual	18
3.2.3. Híbrida: Selección manual y regularización.....	21
3.3. Software y tratamiento previo de datos	22
3.4. Modelos	24
3.4.1. Regresión Logística (LR)	24
3.4.2. Árbol de decisión (DT).....	24
3.4.3. K-vecinos más cercanos (KNN)	24
3.4.4. Bosque aleatorio (RF).....	25
3.4.5. Potenciación del gradiente (GBM)	25
3.4.6. Máquina de Vector de Soporte (SVM).....	25
3.5. Evaluación de desempeño predictivo	26
3.5.1. Balanced accuracy (BAAC)	26
3.5.2. Precisión	26
3.5.3. Recuperación	27
3.5.4. Especificidad	27
3.5.5. F1-Score.....	27
3.5.6. Área bajo la curva ROC (AUC-ROC).....	27
3.6. Análisis de estabilidad en la selección de predictores	27
3.7. Importancia Relativa Ponderada.....	28
4. Resultados.....	30
4.1. Estadísticas descriptivas	30
4.2. Predictores seleccionados	34
4.3. Desempeño predictivo	37
4.4. Análisis de Estabilidad en la selección de predictores	39
4.5. Importancia Relativa Ponderada de los predictores	40

5. Discusión y conclusiones.....	43
6. Limitaciones y recomendaciones.....	47
Referencias	48
Anexos.....	58
Anexo A. Tablas de predictores	58
Anexo B. Tablas de desempeño predictivo	75
Anexo C. Análisis de Estabilidad.....	80
Anexo D. Importancia Relativa Ponderada	88

RESUMEN

Las Tecnologías de la Información y la Comunicación (TIC) se han consolidado como un pilar esencial para la competitividad y el crecimiento económico. El objetivo de este estudio es identificar los factores determinantes de la inversión empresarial en TIC en Ecuador y evaluar la capacidad predictiva de distintos enfoques de selección de predictores y algoritmos de clasificación. Esta investigación analiza la inversión empresarial en TIC en Ecuador a partir de la Encuesta de Estructura Empresarial (ENESEM 2022), que recopiló información de 3,910 empresas de los sectores de manufactura, comercio, servicios, construcción y minería, con representatividad nacional y especial relevancia en polos estratégicos como Pichincha y Guayas. Para abordar el fenómeno se compararon tres estrategias de selección de predictores (regularización, manual e híbrida) en combinación con siete algoritmos de clasificación de aprendizaje automático (ML): Regresión Logística (LR), Árboles de Decisión (DT), K-vecinos más cercanos (KNN), Bosques Aleatorios (RF), Potenciación del Gradiente (GBM) y Máquinas de Vectores de Soporte (SVM) con kernels lineal y radial. También se evaluaron los predictores mediante el análisis de selección de estabilidad y de importancia relativa ponderada. Los resultados muestran que el algoritmo Random Forest logra la mayor precisión, mientras que el SVM lineal también presenta un buen rendimiento. Esto sugiere que los factores predictivos presentan tanto componentes lineales como no lineales en su relación con la inversión en TIC. Se identificó un núcleo estable de variables recurrentes como innovación, especialistas TIC, sitio web, uso de internet para capacitación y comercio electrónico, junto con factores asociados al género, como la presencia de mujeres en cargos administrativos y sus niveles salariales. En conjunto, los hallazgos ofrecen evidencia empírica para comprender la transformación digital empresarial en Ecuador y demuestran el potencial del ML como aporte metodológico en la investigación económica aplicada.

Palabras claves: Tecnología de la información y la comunicación, aprendizaje automático, adopción de las TIC, país en desarrollo, transformación digital, Ecuador, Economía de la innovación.

Clasificación JEL: O32, C63

ÍNDICE DE ILUSTRACIONES

Ilustración 1 - Número de empresas por provincia.....	30
Ilustración 2 - Adopción de TIC por región.	31
Ilustración 3 - Adopción de TIC por sector económico.	33
Ilustración 4 - Importancia ponderada de los 10 principales predictores obtenidos mediante la estrategia de regularización (sin categóricas multinivel, partición = 10).	89
Ilustración 5 - Importancia ponderada de los principales predictores obtenidos mediante la estrategia de regularización (sin categóricas multinivel, partición = 5).	90
Ilustración 6 - Importancia ponderada de los principales predictores obtenidos mediante la estrategia de regularización (con categóricas multinivel, partición = 10).	94
Ilustración 7 - Importancia ponderada de los principales predictores obtenidos mediante la estrategia de regularización (con categóricas multinivel, partición = 5).	97
Ilustración 8 - Importancia ponderada de los principales predictores obtenidos mediante la estrategia manual (con categóricas multinivel, partición = 10).	99
Ilustración 9 - Importancia ponderada de los principales predictores obtenidos mediante la estrategia manual (con categóricas multinivel, partición = 5).	100
Ilustración 10 - Importancia ponderada de los principales predictores obtenidos mediante la estrategia manual (sin categóricas multinivel, partición = 10).	102
Ilustración 11 - Importancia ponderada de los principales predictores obtenidos mediante la estrategia manual (sin categóricas multinivel, partición = 5).	103
Ilustración 12 - Importancia ponderada de los principales predictores obtenidos mediante la estrategia híbrida (con categóricas multinivel, partición = 10).	104
Ilustración 13 - Importancia ponderada de los principales predictores obtenidos mediante la estrategia híbrida (con categóricas multinivel, partición = 5).	106
Ilustración 14 - Importancia ponderada de los principales predictores obtenidos mediante la estrategia híbrida (sin categóricas multinivel, partición = 10).	107
Ilustración 15 - Importancia ponderada de los principales predictores obtenidos mediante la estrategia híbrida (sin categóricas multinivel, partición = 5).	108

ÍNDICE DE TABLAS

Tabla 1 - Predictores identificados mediante la estrategia de selección manual.....	20
Tabla 2 - Predictores identificados mediante la estrategia de selección híbrida.	22
Tabla 3 - Matriz de confusión.....	26
Tabla 4 - Resultados de la prueba de chi-cuadrado para la relación entre la región e inversión en TIC.....	31
Tabla 5 - Distribución de las firmas por tamaño y actividad económica.	32
Tabla 6 - Resultados de la prueba de chi-cuadrado para la relación entre sector económico e inversión en TIC.	33
Tabla 7 - Resultados del Error de Cuadrático Medio (Lasso vs. Group Lasso) por estrategia y número de particiones.....	34
Tabla 8 - Resumen de predictores seleccionados por estrategia y tipo de validación cruzada.	35
Tabla 9 - Predictores con mayor frecuencia de selección en las doce configuraciones.	36
Tabla 10 - Desempeño predictivo por estrategia, variante y tamaño de partición para validación cruzada.	38
Tabla 11 - Top 10 de predictores estables en las estrategias de regularización e híbrida. ..	39
Tabla 12 – Top 15 predictores con mayor importancia relativa ponderada promedio.	42
Tabla 13 - Predictores obtenidos mediante la estrategia de regularización (sin categóricas multinivel, particiones = 10 y 5).....	58
Tabla 14 - Predictores obtenidos mediante la estrategia de regularización (con categóricas multinivel, partición = 10).....	60
Tabla 15 - Predictores obtenidos mediante la estrategia de regularización (con categóricas multinivel, partición = 5).....	65
Tabla 16 - Predictores obtenidos mediante la estrategia híbrida (con categóricas multinivel, partición = 10).	70
Tabla 17 - Predictores obtenidos mediante la estrategia híbrida (con categóricas multinivel, partición = 5).	71
Tabla 18 - Predictores obtenidos mediante la estrategia híbrida (sin categóricas multinivel, partición = 10).	72
Tabla 19 - Predictores obtenidos mediante la estrategia híbrida (sin categóricas multinivel, partición = 5).	73
Tabla 20 - Evaluación de modelos a partir de la estrategia de regularización (sin categóricas multinivel, partición = 10).....	75
Tabla 21 - Evaluación de modelos a partir de la estrategia de regularización (sin categóricas multinivel, partición = 5).....	75
Tabla 22 - Evaluación de modelos a partir de la estrategia de regularización (con categóricas multinivel, partición = 10).....	76
Tabla 23 - Evaluación de modelos a partir de la estrategia de regularización (con categóricas multinivel, partición = 5).....	76
Tabla 24 - Evaluación de modelos a partir de la estrategia de selección manual (con categóricas multinivel, partición = 10).....	77
Tabla 25 - Evaluación de modelos a partir de la estrategia de selección manual (con categóricas multinivel, partición = 5).....	77
Tabla 26 - Evaluación de modelos a partir de la estrategia de selección manual (sin categóricas multinivel, partición = 10).....	77
Tabla 27 - Evaluación de modelos a partir de la estrategia de selección manual (sin categóricas multinivel, partición = 5).....	78

Tabla 28 - Evaluación de modelos a partir de la estrategia híbrida (con categóricas multinivel, partición = 10).	78
Tabla 29 - Evaluación de modelos a partir de la estrategia híbrida (con categóricas multinivel, partición = 5).	78
Tabla 30 - Evaluación de modelos a partir de la estrategia híbrida (sin categóricas multinivel, partición = 10).	79
Tabla 31 - Evaluación de modelos a partir de la estrategia híbrida (sin categóricas multinivel, partición = 5).	79
Tabla 32 - Predictores estables de la estrategia de regularización (sin categóricas multinivel, particiones = 10 y 5).	80
Tabla 33 - Predictores estables de la estrategia de regularización (con categóricas multinivel, partición = 10).	81
Tabla 34 - Predictores estables de la estrategia de regularización (con categóricas multinivel, partición = 5).	83
Tabla 35 - Predictores estables de la estrategia híbrida (con categóricas multinivel, partición = 10).	85
Tabla 36 - Predictores estables de la estrategia híbrida (con categóricas multinivel, partición = 5).	85
Tabla 37 - Predictores estables de la estrategia híbrida (sin categóricas multinivel, partición = 10).	86
Tabla 38 - Predictores estables de la estrategia híbrida (sin categóricas multinivel, partición = 5).	87
Tabla 39 - Importancia ponderada de los predictores obtenidos mediante la estrategia de regularización (sin categóricas multinivel, partición = 10).	88
Tabla 40 - Importancia ponderada de los predictores obtenidos mediante la estrategia de regularización (sin categóricas multinivel, partición = 5).	89
Tabla 41 - Importancia ponderada de los predictores obtenidos mediante la estrategia de regularización (con categóricas multinivel, partición = 10).	91
Tabla 42 - Importancia ponderada de los predictores obtenidos mediante la estrategia de regularización (con categóricas multinivel, partición = 5).	95
Tabla 43 - Importancia ponderada de los predictores obtenidos mediante la estrategia manual (con categóricas multinivel, partición = 10).	98
Tabla 44 - Importancia ponderada de los predictores obtenidos mediante la estrategia manual (con categóricas multinivel, partición = 5).	99
Tabla 45 - Importancia ponderada de los predictores obtenidos mediante la estrategia manual (sin categóricas multinivel, partición = 10).	101
Tabla 46 - Importancia ponderada de los predictores obtenidos mediante la estrategia manual (sin categóricas multinivel, partición = 5).	102
Tabla 47 - Importancia ponderada de los predictores obtenidos mediante la estrategia híbrida (con categóricas multinivel, partición = 10).	103
Tabla 48 - Importancia ponderada de los predictores obtenidos mediante la estrategia híbrida (con categóricas multinivel, partición = 5).	105
Tabla 49 - Importancia ponderada de los predictores obtenidos mediante la estrategia híbrida (sin categóricas multinivel, partición = 10).	106
Tabla 50 - Importancia ponderada de los predictores obtenidos mediante la estrategia híbrida (sin categóricas multinivel, partición = 5).	107

ABREVIATURAS

TIC	Tecnologías de la Información y la Comunicación
ENESEM	Encuesta de Estructura Empresarial
INEC	Instituto Nacional de Estadísticas y Censos
ML	Aprendizaje Automático
LASSO	Least Absolute Shrinkage and Selection Operator
GLASSO	Group LASSO
CV	Validación Cruzada (Cross Validation)
ECM	Error Cuadrático Medio
LR	Regresión Logística
DT	Árbol de decisión
KNN	K-vecinos más cercanos
RF	Bosque aleatorio
GBM	Potenciación del gradiente
SVM	Máquina de vector de soporte
ROSE	Random Over-Sampling Examples
I+D	Investigación y Desarrollo
BACC	Balanced Accuracy
AUC	Área bajo la curva ROC

1. INTRODUCCIÓN

En la era digital, las Tecnologías de la Información y la Comunicación (TIC) se han convertido en pilares de la vida cotidiana y del ámbito empresarial, al transformar la educación, la salud, el comercio y el trabajo, e impulsar la competitividad y el crecimiento económico (Osunmadewa et al., 2024). De esta forma, estas tecnologías han cambiado la forma en que los individuos interactúan con el mundo, permitiéndoles aprovechar posibilidades antes inalcanzables al permitir romper las fronteras geográficas y aprovechar estos recursos de comunicación (Bolliger & Halupa, 2018; H. Wang et al., 2023). De hecho, ITU (2025), que es el organismo especializado de las Naciones Unidas para las TIC estimó que aproximadamente el 64% de la población mundial (5.1 mil millones) utilizaron internet en 2022 y esta cifra aumenta al 68% (5.5 mil millones) en 2024.

Más allá de su impacto en la vida cotidiana de las personas, las TIC también desempeñan un papel transformador en el ámbito empresarial puesto que son un factor clave para aumentar la eficiencia (Piabuo et al., 2017), la competitividad (Cuevas-Vargas et al., 2015), la productividad (Díaz-Chao et al., 2015; Luo & Bu, 2016) y la internacionalización de las firmas (Hagsten & Kotnik, 2017). Específicamente, las TIC permiten a las organizaciones reaccionar rápidamente a los cambios del mercado, optimizar los procesos mediante la comunicación y cooperación en tiempo real entre equipos de todo el mundo, con tecnologías como la videoconferencia y mensajería instantánea (Achanta, 2024; Belfo et al., 2015; Osunmadewa et al., 2024), lo que permite agilizar procesos, reducir costos y acceder a información en tiempo real para la toma de decisiones.

No obstante, a pesar de la evidente importancia de las TIC en el entorno empresarial, sigue siendo un desafío comprender a profundidad los factores que impulsan su adopción y las decisiones estratégicas relacionadas con su inversión. Estudios previos han aplicado técnicas de aprendizaje automático (ML, por sus siglas en inglés) para predecir variables relacionadas con el uso de TIC y la innovación empresarial. Por ejemplo, Kumari & Singh (2022) utiliza modelos avanzados para pronosticar el índice de uso de TIC y el PIB, mientras que Eom et al. (2024) predice el rendimiento de la innovación en procesos de negocio relacionados con TIC. Sin embargo, estos trabajos se centran en contextos específicos, dejando un vacío en la literatura respecto a la aplicación de modelos predictivos para entender la inversión empresarial en TIC, un aspecto clave para evaluar la digitalización a nivel organizacional. Este estudio llena esa brecha y se enfoca en la predicción de una fase previa crítica: la decisión de invertir en TIC.

Este estudio compara tres enfoques de selección de predictores (regularización, manual e híbrido) aplicados a siete modelos de aprendizaje automático, con el fin de evaluar su capacidad predictiva y la estabilidad de los determinantes de la inversión en TIC. Posteriormente, se contrastan estas tres estrategias mediante siete modelos de aprendizaje automático: Regresión Logística (LR), Árbol de decisión (DT), K-vecinos más cercanos (KNN), Bosque aleatorio (RF), Potenciación del gradiente (GBM), y Máquina de vector de soporte (SVM) con kernels lineal y radial. De esta manera, este diseño metodológico permite no solo comparar el desempeño predictivo de distintos algoritmos, sino también analizar la estabilidad de los determinantes de la inversión en TIC en función del método de selección de predictores.

Esta investigación aporta en dos planos. En el empírico, identifica patrones de adopción tecnológica a partir de microdatos empresariales y modelos de aprendizaje automático. En el metodológico, contrasta distintas técnicas de selección de variables y algoritmos de clasificación, lo que permite evaluar la estabilidad de los resultados y el potencial predictivo de cada enfoque. Estos hallazgos pueden servir de insumo tanto para la literatura académica en economía de la empresa como para el diseño de políticas orientadas a promover la transformación digital.

El resto del documento se organiza de la siguiente manera. El capítulo 2 presenta la revisión de literatura; el capítulo 3 describe los datos, variables y modelos; el capítulo 4 expone los resultados; el capítulo 5 discute los hallazgos, y el capítulo 6 plantea las limitaciones y futuras líneas de investigación.

2. REVISIÓN LITERARIA

Esta sección revisa estudios relevantes sobre la influencia de las TIC en el entorno empresarial y el uso de técnicas de aprendizaje automático para anticipar su adopción, señalando las lagunas existentes que esta investigación pretende abordar.

2.1. Relevancia de las TIC en las empresas

Las Tecnologías de la Información y la Comunicación (TIC) han transformado la productividad, la innovación y la competitividad de las empresas en distintos sectores. Se han convertido en un componente central de las estrategias empresariales modernas, al facilitar la interacción con clientes y la automatización de procesos que incrementan la eficiencia (Díaz-Chao et al., 2015; Luo & Bu, 2016; Osunmadewa et al., 2024). Además, promueven la colaboración entre trabajadores mediante plataformas digitales, fortaleciendo las competencias de trabajo en equipo (Almerich et al., 2021; Lewin & McNicol, 2015), y el uso de grandes volúmenes de datos (en inglés, *big data*) ofrece a las organizaciones la posibilidad de tomar decisiones mejor fundamentadas (Gheisari, 2024).

La transformación digital también impulsa nuevos modelos de negocio que otorgan a las empresas una ventaja competitiva sostenible (Xue et al., 2022). Asimismo, fomenta la innovación mediante la colaboración y la co-creación con los clientes, quienes participan en el diseño y la retroalimentación de productos y procesos (Chepurna & Rialp Criado, 2018; Patro & Raghunath, 2021). En conjunto, las empresas que integran eficazmente las TIC pueden adaptarse mejor a los cambios del mercado y sostener altos niveles de rendimiento y competitividad (Yunis et al., 2018).

2.2. Antecedentes en la región latinoamericana

En el contexto de América Latina, se han publicado escasos estudios acerca de las TIC. No obstante, se han obtenido resultados interesantes con respecto a su incidencia en distintas cuestiones empresariales. En un contexto de gestión de procesos, Gomez-Sanchez et al. (2023) encontraron que las TIC tienen un impacto fuerte y directo en las decisiones de importación de las empresas del sector manufacturero colombiano, especialmente en las PYME. Asimismo, Grazzi & Jung (2019) hallaron la presencia de efectos epidémicos y de rango, en el sentido de que hay una mayor probabilidad de que las firmas más grandes antiguas, intensivas en mano de obra cualificada, exportadoras y urbanas adopten las TIC; no obstante, una vez adoptadas, el tamaño y la ubicación pierden importancia. En el ámbito

de los recursos humanos, las TIC han posibilitado el teletrabajo, sin embargo, evidencia proveniente de Colombia, Brasil y Argentina, muestra que, aunque la pandemia COVID-19 incrementó el teletrabajo, su uso no ha sido masivo en estas regiones (Oviedo Gil & Cala Vitery, 2023).

Adicionalmente, en un nivel macroeconómico, Rabanal (2024) analizó la paradoja de la productividad y halló que el uso de internet (como proxy de TIC) por parte de las familias tiene una contribución positiva pero inelástica sobre la productividad de la economía. Mientras tanto, en materia de política económica y pública, Finkelstein Shapiro et al. (2024) encontraron que una política que reduce los costos de adopción digital por parte de las empresas impulsa la recuperación del PIB, el empleo y los ingresos laborales tras la COVID-19 en América Latina. Por su parte, Alvarado Vazquez et al. (2023) realizaron un estudio en México, sobre el uso de TIC para la participación social en la planificación, el diseño y el mantenimiento de espacios públicos, y evidenciaron una brecha entre la expectativa de que las TIC fortalezcan la participación ciudadana y la realidad institucional, donde predominan las restricciones presupuestarias, apoyo político y de capacidades técnicas.

2.3. Uso de técnicas de aprendizaje automático en la investigación sobre las TIC

El avance de las TIC ha impulsado el uso de métodos de aprendizaje automático (ML, por sus siglas en inglés) como alternativa a los enfoques tradicionales, ya que permiten analizar grandes volúmenes de datos, identificar patrones no lineales y elaborar predicciones con alta precisión (James et al., 2023; Zhou, 2021). Por ejemplo, Verma & Illes (2019) presentaron un estudio experimental que se centró en predecir la actitud de estudiantes hacia las TIC y la tecnología móvil en universidades de India y Hungría, aplicando varios algoritmos de clasificación, entre ellos redes neuronales y máquinas de soporte vectorial, y encontró que el modelo SVM alcanzó el mejor desempeño en ambos contextos. Este tipo de investigación no sólo aumenta nuestra comprensión de la forma en que afectan las TIC a la educación, sino que también demuestra lo prometedores que son los modelos predictivos para implantar sistemas de apoyo en tiempo real.

En el ámbito empresarial, un estudio reciente realizado por Eom et al. (2024) aplicaron ML para predecir el rendimiento de la innovación en procesos de negocio, y hallaron que el Bosque Aleatorio fue el modelo más efectivo. Además, identificaron que las fuentes de información constituyen el factor más crítico para explicar el desempeño

innovador. Este enfoque destaca la forma en que los métodos de aprendizaje automático pueden proporcionar información útil para que los líderes empresariales optimicen sus estrategias de innovación.

Por otro lado, la investigación de Herrera et al. (2023) sobre el impacto de las TIC en las características socioeconómicas incorporó enfoques de aprendizaje automático explicable, combinando modelos como XGBoost y redes neuronales con valores SHAP, lo que permitió identificar relaciones complejas entre TIC, educación e ingresos, además de resaltar la importancia de políticas públicas enfocadas en equidad de género e inclusión digital.

Estos ejemplos muestran que el ML en la investigación sobre TIC permite abordar problemas complejos, identificar factores clave y desarrollar herramientas predictivas. No obstante, persisten desafíos en la selección de variables y en la interpretación de los modelos, aspectos que deben fortalecerse para asegurar resultados robustos y aplicables.

2.4. Brecha en la literatura

En las últimas dos décadas, la literatura sobre adopción y uso de TIC en empresas ha crecido de manera significativa, con énfasis en la gestión del conocimiento (Dave & Koskela, 2009; García-Álvarez, 2015; Rocha et al., 2024), la innovación (Eom et al., 2024; Hall et al., 2013; S. Wang & Zhang, 2025) y los efectos en la productividad (Cardona et al., 2013; Hall et al., 2013). Sin embargo, son escasas las investigaciones que analizan de forma directa la decisión de invertir en TIC y, en particular, aquellas que lo hacen a partir de microdatos empresariales y con enfoques predictivos. Esta limitación restringe la comprensión de los elementos que explican la destinación efectiva de recursos a la construcción de capacidades digitales.

Este estudio contribuye a cerrar esta brecha al aplicar modelos de aprendizaje automático junto con técnicas rigurosas de selección de predictores, como LASSO y Group LASSO, sobre microdatos empresariales representativos a nivel nacional. Este enfoque no solo mejora la capacidad predictiva, sino que también permite identificar de manera empírica los determinantes económicos de la inversión en TIC. La contribución resulta especialmente pertinente en el contexto de países en desarrollo, donde este tipo de análisis sigue siendo limitado. De este modo, la investigación vincula explícitamente la predicción cuantitativa de la inversión con el análisis de sus determinantes económicos, ofreciendo evidencia robusta y novedosa para el caso ecuatoriano.

3. METODOLOGÍA

3.1. Datos

En este apartado se describe la construcción de la base de datos utilizada para la investigación, las consideraciones sobre ponderación y representatividad, así como la definición de la variable de interés empleada en los modelos.

3.1.1. *Construcción de la base*

Este estudio utiliza datos de la Encuesta de Estructura Empresarial (ENESEM 2022), levantada por el Instituto Nacional de Estadísticas y Censos (INEC) en Ecuador. La base incluyó inicialmente información de 3,951 empresas y, tras el proceso de depuración y limpieza, se obtuvieron 3,910 observaciones válidas. Cada observación en la muestra cuenta con su propio factor de expansión, cuyo valor varía según el estrato de diseño muestral. Estos factores fluctúan entre 1 y 498.52, con un valor promedio de 135.07. La suma de todos ellos expande la muestra a una población de alrededor de 15,110 empresas.

La encuesta se organiza en dos secciones principales. El Tomo 1 incluye 646 variables relacionadas con agregados económicos como producción, consumo intermedio y valor agregado en los sectores de manufactura, minería, comercio, construcción y servicios. La sección TIC incorpora 141 variables vinculadas con inversión tecnológica, conexión y uso de internet, comercio electrónico y herramientas digitales (INEC, 2023).

3.1.2. *Estrategia de ponderación y representatividad*

En este estudio, los factores de expansión se aplicaron únicamente en la etapa de estadísticas descriptivas. Esto permite que porcentajes, promedios e indicadores reflejen de manera representativa a la población total de empresas y no solo a la muestra encuestada, garantizando comparabilidad con cifras oficiales y evitando sesgos del diseño muestral.

En contraste, los modelos predictivos se estimaron sin ponderación, ya que el objetivo no era la inferencia poblacional, sino comparar el desempeño de distintos métodos de selección de predictores y algoritmos. Esta decisión responde a la necesidad de preservar la estabilidad numérica y se ajusta a la práctica consolidada en la literatura, donde los pesos muestrales rara vez se incorporan en la modelación predictiva (J. Friedman, 2013; Winship & Radbill, 1994). Estudios recientes en aprendizaje automático con datos de encuestas señalan que, si bien los pesos pueden mejorar la evaluación poblacional de métricas, en la fase de entrenamiento suelen emplearse alternativas como el balanceo de clases (Wadekar & Reiter, 2024).

3.1.3. *Variable de interés*

La variable dependiente de este estudio es la inversión en TIC. Se deriva del cuestionario ENESEM 2022, que pregunta explícitamente a las empresas: "¿Ha invertido su empresa en tecnologías de la información y la comunicación (TIC) durante el año 2022?". La respuesta es binaria, con valores codificados como 1 si la empresa declaró haber invertido en TIC y 0 en caso contrario. No se proporcionaron opciones de respuesta adicionales en la encuesta.

3.2. Selección de predictores

En primer lugar, se identificaron y eliminaron las variables redundantes que eran comunes entre las bases de datos utilizadas, obteniéndose un conjunto inicial de 759 variables. Posteriormente, tras aplicar procedimientos de depuración (tratamiento de valores inconsistentes y eliminación de variables con más del 90% de datos faltantes), así como la agrupación de indicadores derivados, se obtuvo un total de 739 variables. La elevada dimensionalidad de este conjunto justifica la incorporación de técnicas de regularización y selección de predictores, con el fin de mitigar problemas de sobreajuste y multicolinealidad, y mejorar la eficiencia computacional de los modelos de predicción. Por ende, para abordar este desafío, se implementaron tres estrategias:

3.2.1. *Regularización*

La primera estrategia consistió en aplicar técnicas de regularización, específicamente Ridge y LASSO (Least Absolute Shrinkage and Selection Operator), para identificar las variables más relevantes. Estas metodologías no sólo permiten controlar el sobreajuste, sino que también facilitan la identificación de las variables más influyentes para la predicción de la inversión en TIC (James et al., 2023). Mientras que Ridge penaliza los coeficientes grandes para mitigar problemas de multicolinealidad, LASSO se destaca al asignar coeficientes cero a las variables menos significativas, permitiendo una selección más efectiva de predictores claves (Fonti & Belitser, 2017; James et al., 2023; Osman et al., 2017). Es importante recalcar que existe, una extensión de LASSO, llamada Group LASSO, donde se considera el caso en que hay variables ficticias en grupo, es decir, que están en formato one-hot encoding.

Para esta estrategia, se partió de la base completa de variables, es decir, 738 predictores para la variable de interés. La estrategia se aplicó en dos variantes:

1. Sin categóricas multinivel: Se conservaron únicamente las variables binarias y numéricas; excluyendo las variables categóricas con más de dos niveles, debido a que su transformación en múltiples variables ficticias podía generar redundancias.
2. Con categóricas multinivel: Se incorporaron las variables categóricas con más de dos niveles, las cuales fueron transformadas en variables ficticias mediante one-hot encoding.

3.2.2. Selección Manual

La segunda estrategia se basó en la selección de predictores respaldados por la literatura sobre adopción tecnológica y economía de la innovación. Se identificaron 19 variables agrupadas en cuatro bloques temáticos: (i) características de la empresa, (ii) desempeño financiero, (iii) recursos humanos, y (iv) innovación y TIC. Esta estructura permitió mantener variables con fundamento teórico, incluso cuando no hubieran sido priorizadas por métodos automatizados. La Tabla 1 resume las variables seleccionadas y sus principales referencias.

Características de la empresa

Las características estructurales de la empresa constituyen un determinante clásico de la adopción tecnológica. En primer lugar, con respecto a la antigüedad de la firma existe dos visiones distintas, ya que, se asocia con la experiencia acumulada y con ello, la capacidad organizacional para implementar cambios; no obstante, también puede reflejar rigidez para adaptarse a los constantes desarrollos externos (Ben Youssef et al., 2011; Sørensen & Stuart, 2000). Por otra parte, la localización geográfica influye en la inversión de TIC a través de los efectos de aglomeración, ya que, hay territorios que atraen inversión debido a que facilitan el acceso a infraestructura digital y a capital humano especializado (Nepelski & De Prato, 2014; Vas et al., 2024). Asimismo, la pertenencia sectorial determina la intensidad de la competencia y la presión tecnológica, incluso, se ha demostrado que las empresas que se encuentran en sectores donde hay mayores oportunidad de crecimiento invierten más en TIC (Neirotti & Pesce, 2019).

Desempeño financiero

El desempeño financiero condiciona directamente la capacidad de las empresas para invertir en nuevas tecnologías. Por una parte, las ventas y el valor agregado empresarial son indicadores de tamaño económico y la productividad de las firmas, variables positivamente asociadas con la adopción tecnológica (Duc & Nguyen, 2023; Koutroumpis

et al., 2020). Al mismo tiempo, la estructura financiera, medida a través del endeudamiento sobre patrimonio, refleja restricciones o facilidades para financiar proyectos de innovación (Nukala & Prasada Rao, 2021; L. Zhang et al., 2019). De esta forma, las variables financieras permiten capturar tanto la disponibilidad de recursos como las limitaciones que enfrenta la empresa en sus decisiones de inversión digital.

Recursos humanos

Otro factor ampliamente documentado en la literatura sobre la adopción de TIC e innovación es la composición del capital humano (Gimenez & Vargas-Montoya, 2021). Por un lado, una primera variable utilizada como proxy del tamaño organizacional es el total de personal afiliado que posee la firma, el cual se asocia a una mayor capacidad de absorción de conocimiento y tecnología (Danquah & Amankwah-Amoah, 2017; G. Lee & Xia, 2006). La presencia de profesionales científicos e intelectuales refleja un capital humano altamente calificado en cuanto a habilidades y conocimientos, lo cual es fundamental para la generación y asimilación de nuevas tecnologías en las empresas (Giotopoulos et al., 2017; Guerrieri et al., 2011; Molla et al., 2024; Murphy & Siedschlag, 2013). Finalmente, se incorpora la dimensión de género, aproximada mediante el número de mujeres en la fuerza laboral, dado que la literatura ha documentado diferencias en los patrones de inversión y adopción tecnológica según el sexo (Anggraini et al., 2023; Mehta & Sinha, 2022; Orser & Riding, 2018; Živković et al., 2024).

Innovación y TIC

Las variables de innovación y TIC representan directamente las capacidades tecnológicas de la empresa. Las diversas fuentes de Innovación y Desarrollo (I+D) estimulan el intercambio y la adquisición de conocimiento (Giotopoulos et al., 2017; S. Lee et al., 2016), por ello, se consideró las variables de I+D (interna y externa), así como, la existencia de un departamento específico de I+D. En esta misma línea, la presencia de especialistas TIC refleja la disponibilidad de competencias digitales dentro de la organización, así como, personal capacitado para identificar soluciones tecnológicas pertinentes al negocio, lo que a su vez, facilita la toma de decisiones fundamentadas respecto a la adopción (Eze & Chinedu-Eze, 2018; Eze & Vera, 2018). En paralelo, las variables de infraestructura digital (conexión a internet, sitio web, firma digital) y de comercio electrónico (ventas y compras por internet) reflejan el grado de digitalización de las operaciones de la empresa (Giotopoulos et al., 2017). En conjunto, este bloque capta tanto la capacidad tecnológica como la orientación innovadora que explican la propensión a invertir en TIC.

Tabla 1 - Predictores identificados mediante la estrategia de selección manual.

Bloque	Variables	Descripción	Fundamento teórico
Características de la empresa	Edad	Edad de la empresa	Ben Youssef et al. (2011); Sørensen & Stuart (2000)
	Región (Costa, Amazonía, Insular)*	Localización geográfica	Nepelski & De Prato (2014); Vas et al. (2024)
	Sector (Minería, Comercio, Construcción, Servicios)*	Pertenencia sectorial	Neirotti & Pesce (2019)
Desempeño Financiero	Ventas	Ventas totales	Duc & Nguyen (2023)
	VAE	Valor agregado empresarial	Koutroumpis et al. (2020)
	Endeudamiento/Patrimonio	Ratio de endeudamiento	Nukala & Rao (2021); Zhang et al. (2019)
Recursos Humanos	Total_PA	Total de personal afiliado	Danquah & Amankwah-Amoah (2017)
	PA_Científicos	Profesionales científicos e intelectuales	Giotopoulos et al. (2017)
	Total_PA_Mujeres	Mujeres en la fuerza laboral	Anggraini et al. (Anggraini et al., 2023); Orser & Riding (2018)
Innovación y TIC	Innovación	Actividades de I+D	Giotopoulos et al. (2017); Lee et al. (2016)
	PA_Especialista_TIC	Presencia de especialistas TIC	Eze & Chinedu-Eze (2018)
	I+D Interna Externa Departamento	Inversión y estructura en I+D	Guerrieri et al. (2011); Molla et al. (2024)
	Internet Sitio web Firma digital	Infraestructura digital	Giotopoulos et al. (2017)
	Ventas y compras por internet	Comercio electrónico	Eze & Vera (2018)

* Estas son las variables que se excluyeron para la segunda variante (sin variables categóricas multinivel).

Por otro lado, es importante mencionar que si bien algunas de estas variables presentan posibles problemas de multicolinealidad; se optó por conservarlas dado que cada una cuenta con un respaldo teórico específico. Asimismo, en este sentido, el riesgo de multicolinealidad no invalida el enfoque, ya que el objetivo principal es la predicción y no la estimación causal de coeficientes, y además los modelos de aprendizaje automático empleados (árboles, bosques aleatorios, SVM, entre otros) son menos sensibles a este problema.

El objetivo fue contrastar la capacidad predictiva de un conjunto reducido y fundamentado de variables frente a la selección automática del LASSO. La estrategia se aplicó en dos variantes:

1. Con categóricas multinivel: Se mantuvieron las variables categóricas con más de dos niveles que proveen información sobre el sector económico y la región donde se encuentra la empresa. Estas variables fueron transformadas en variables ficticias mediante one-hot encoding.
2. Sin categóricas multinivel: Se excluyeron las variables categóricas con más de dos niveles. Esta variante permitió evaluar si la inclusión de sector y región aportaba capacidad predictiva adicional o si, por el contrario, los resultados se mantenían estables sin ellas.

3.2.3. Híbrida: Selección manual y regularización

La estrategia híbrida combinó la fundamentación teórica de la selección manual con la parsimonia de la regularización. Se partió de un conjunto ampliado de predictores en cada bloque temático (características de la empresa, desempeño financiero, recursos humanos e innovación y TIC), que fue sometido a LASSO o GLASSO según correspondiera. Este enfoque permitió depurar un conjunto inicial de 33 variables en función de su capacidad explicativa, aprovechando la lógica conceptual de los bloques temáticos y la eficiencia de la regularización automática (Tabla 2). Posteriormente se aplicó LASSO o GLASSO según correspondiera. Se definieron dos variantes:

1. Con categóricas multinivel: El conjunto manual ampliado incluyendo las variables de sector económico y región (codificados en variables ficticias). Dado este tratamiento, se aplicó GLASSO para respetar la estructura en grupos.
2. Sin categóricas multinivel: En esta versión se eliminaron dichas categóricas, por lo que se aplicó LASSO sobre un conjunto exclusivamente numérico y binario.

Tabla 2 - Predictores identificados mediante la estrategia de selección híbrida.

No.	Bloque temático	Cantidad	Variables
1	Características de la empresa	3	Edad, región* y sector económico*.
2	Desempeño financiero	11	Ventas, ingresos, costos, gastos, nivel de endeudamiento (activo y patrimonio), valor agregado empresarial, retorno (sobre la inversión y el patrimonio), margen neto, y formación de capital fijo empresarial.
3	Recursos humanos	7	Personal afiliado total, profesionales científicos e intelectuales, mujeres, directores y gerentes, y especialistas TIC (total, hombre, mujer).
4	Innovación y TIC	12	Innovación, I+D (interno y externo), departamento específico de I+D, conexión a internet, ventas y compras por internet, disposición de sitio web y firma digital, personal afiliado que usa internet (hombres y mujeres) y total de dispositivos.

* Estas son las variables que se excluyeron para la segunda variante (sin variables categóricas multinivel)

De esta forma, se busca aprovechar el conocimiento teórico en la elección de predictores, al mismo tiempo que se emplea un enfoque automatizado para reducir la dimensionalidad y mejorar la capacidad predictiva de los modelos. En consecuencia, la estrategia híbrida permitió contrastar si la regularización, aplicada sobre un conjunto previamente delimitado, lograba una mayor parsimonia y robustez en los predictores seleccionados.

3.3. Software y tratamiento previo de datos

Este apartado describe el preprocesamiento y la partición de los datos, con el objetivo de asegurar claridad, reproducibilidad y generalización de los modelos, reduciendo el riesgo de sobreajuste y permitiendo una evaluación precisa en datos reales.

El preprocesamiento se realizó en Python 3.11.9 y el desarrollo de los modelos en R 4.4.1, combinando la eficiencia de Python en la manipulación de datos con la robustez de R en el análisis estadístico y de ML. Las estadísticas descriptivas con factores de expansión se elaboraron en IBM SPSS Statistics 31.0.0.0, cuya interfaz facilita la ponderación de casos.

Para garantizar la reproducibilidad, se fijó una semilla de 48, asegurando consistencia en las particiones y en los procesos de entrenamiento y validación. Los datos se dividieron en 70% para entrenamiento y 30% para prueba, siguiendo un estándar en aprendizaje automático que equilibra el tamaño del conjunto de ajuste con el de evaluación (Nguyen et al., 2021).

Durante el entrenamiento se aplicó validación cruzada anidada con 10 particiones, siguiendo a Kohavi (1995), lo que permitió estimar de manera robusta el desempeño en datos no usados durante el ajuste y equilibrar sesgo y varianza en el error. Para verificar la estabilidad, el procedimiento se replicó con 5 particiones, contrastando así la robustez de las métricas en distintas configuraciones.

Un desafío común en el entrenamiento de clasificadores de ML es el desbalance en la variable objetivo, lo que puede incrementar el sobreajuste y afectar la capacidad de generalización (Selamat et al., 2022). Por lo tanto, se aplicó la técnica ROSE (Random Over-Sampling Examples) propuesta por Menardi & Torelli (2014), puesto que, la literatura ha demostrado que tiene un buen rendimiento cuando se trata de conjuntos de datos desequilibrados (Demir & Şahin, 2022). Específicamente, ROSE se caracteriza por producir instancias sintéticas en clases minoritarias simulando observaciones que corresponden a la distribución original de los datos (Menardi & Torelli, 2014). De esta forma, la aplicación ROSE pretende equilibrar las clases potenciando la capacidad de los modelos para aprender patrones representativos en ambas categorías y aumentar la capacidad de generalización de las predicciones en circunstancias reales.

Para la clasificación binaria se empleó el criterio de Youden (1950), que selecciona el umbral de decisión que maximiza simultáneamente la sensibilidad y la especificidad (Berrar, 2019). Esto evita la arbitrariedad de usar el valor convencional de 0.5 y ajusta el punto de corte según las características empíricas de los datos, reduciendo sesgos derivados del desbalance (Kuhn & Johnson, 2013).

Finalmente, para asegurar comparabilidad entre modelos y evitar sesgos por magnitud de las variables, se aplicó de manera uniforme la normalización Z-score a todos los predictores. Este procedimiento es esencial en algoritmos sensibles a la escala como

SVM, KNN y regresión logística, y se extendió al resto de modelos para mantener consistencia metodológica (Henderi, 2021; Singh & Singh, 2020).

3.4. Modelos

3.4.1. Regresión Logística (LR)

La regresión logística (en inglés: *Logistic Regression (LR)*) desarrollado por Cox (1958) es un algoritmo de clasificación ampliamente utilizado para predecir variables categóricas binarias (Chang et al., 2024). Este modelo estima la probabilidad de que la variable de respuesta pertenezca a una categoría particular en función de un conjunto de variables independientes (James et al., 2023). La regresión logística transforma una combinación lineal de predictores en una probabilidad mediante la función sigmoide, lo que permite clasificar observaciones en dos categorías (Zaidi, 2022). Se caracteriza por su eficiencia, rapidez de entrenamiento y facilidad de implementación (Chang et al., 2024).

3.4.2. Árbol de decisión (DT)

Los clasificadores de árboles de decisión (en inglés: *Decision Tree (DT)*) son uno de los métodos más conocidos para la representación de clasificación, dado que son algoritmos no paramétricos que particionan iterativamente el espacio de datos en subconjuntos homogéneos según las características predictoras (Czajkowski & Kretowski, 2019; Jijo & Abdulazeez, 2021; H. Zhang et al., 2014). De esta forma, estos algoritmos son atractivos por su facilidad de uso, velocidad de clasificación y efectividad, además, su estructura jerárquica es bastante similar al proceso humano de toma de decisiones (Czajkowski & Kretowski, 2019). Los árboles de decisión dividen el espacio de datos en subconjuntos homogéneos mediante reglas jerárquicas basadas en predictores, lo que los hace intuitivos y comparables al proceso humano de toma de decisiones (Navada et al., 2011).

3.4.3. K-vecinos más cercanos (KNN)

La clasificación del vecino más próximo, también conocida en inglés como *K-nearest neighbors (KNN)* basada usualmente en la distancia euclídeana, se fundamenta en la premisa de que los patrones más próximos a un patrón objetivo proporcionan información útil sobre la etiqueta (Kramer, 2013). De esta forma, KNN asigna etiquetas de clase a una observación en función de las clases predominantes entre sus k vecinos más cercanos, por tal razón, es necesario definir una medida de similitud en el espacio de datos (Chang et al., 2024; Kramer, 2013). Para esta investigación, la medida k es 5. Este algoritmo es sencillo de

implementar, aunque su desempeño depende de la escala de las variables, por lo que requiere normalización previa para evitar distorsiones.

3.4.4. Bosque aleatorio (RF)

Bosque aleatorio (en inglés: *Random Forest (RF)*) desarrollado por Breiman (2001), es un refinamiento adicional de los modelos de árbol de clasificación y regresión (CART). En este modelo, múltiples árboles se entrenan en subconjuntos de datos y la predicción final se obtiene por mayoría de votos. Esto lo hace robusto frente a ruido y capaz de manejar datos faltantes y desbalanceados (Chang et al., 2024; Rigatti, 2017; Tripoliti et al., 2013).

3.4.5. Potenciación del gradiente (GBM)

Potenciación del gradiente (en inglés: *Gradient Boosting Machine (GBM)*) que originalmente fue desarrollado por Friedman (2001) son una clase de sofisticadas técnicas de aprendizaje automático que han sido muy útiles en diversas aplicaciones prácticas (Natekin & Knoll, 2013). Específicamente, el GBM construye árboles de decisión de forma secuencial, corrigiendo en cada paso los errores del modelo anterior para mejorar progresivamente la precisión (Chang et al., 2024). Por lo tanto, en cada paso, el método añade de forma iterativa un nuevo árbol de decisión que disminuye al máximo la función de pérdida (Touzani et al., 2018).

3.4.6. Máquina de Vector de Soporte (SVM)

Otra de las técnicas clásicas de clasificación en aprendizaje automático es Máquina de Vector de Soporte (en inglés: *Support Vector Machine (SVM)*) desarrollada por Vapnik et al. (1996). Este método se caracteriza por encontrar un hiperplano que separe las clases de datos de manera óptima, maximizando la distancia entre las observaciones más cercanas de cada clase, también llamadas vectores de soporte (Suthaharan, 2016). En esta técnica, cuando los datos no son linealmente separables, se emplean funciones de núcleo para transformarlos en un espacio de mayor dimensión, lo que permite la separación lineal (Ghosh et al., 2019). Este método de utilización de funciones puede ser a través de kernels: lineal, polinómicos, función de base radial (RBF, por sus siglas en inglés) o sigmoidales (Meyer & Wien, 2001). En este estudio se emplearon dos kernels: el lineal, adecuado para datos con alta dimensionalidad y con problemas linealmente separables, y el radial (RBF), útil para capturar relaciones no lineales complejas (Ghosh et al., 2019).

3.5. Evaluación de desempeño predictivo

La matriz de confusión es una herramienta clave para examinar el rendimiento de los modelos de clasificación, sobre todo cuando las clases están desequilibradas.

Esta matriz clasifica los resultados de las predicciones en cuatro categorías (véase Tabla 3): Verdaderos Positivos (VP) y Verdaderos Negativos (VN), cuando el modelo acierta en la clase positiva o negativa, respectivamente; y Falsos Positivos (FP) y Falsos Negativos (FN), cuando se equivoca asignando la clase positiva o negativa.

Tabla 3 - Matriz de confusión.

	Valor real positivo (1)	Valor real negativo (0)
Valor predicho positivo (1)	Verdadero Positivo (VP)	Falso Positivo (FP)
Valor predicho negativo (0)	Falso Negativo (FN)	Verdadero Negativo (VN)

A partir de estas categorías, se calculan varias métricas clave que permiten analizar el rendimiento del modelo desde diferentes perspectivas:

3.5.1. *Balanced accuracy (BAAC)*

Como el conjunto de entrenamiento está balanceado (a través de la técnica ROSE), pero el de prueba no, se utiliza la métrica de *balanced accuracy*, dado que promedia la sensibilidad y la especificidad, permitiendo una evaluación más equitativa del rendimiento del modelo (Berrar, 2019). Cabe mencionar que, se optó por dejar el nombre de esta métrica en su idioma original (inglés) para evitar confusiones con la traducción de la métrica de precisión.

$$\text{Balanced accuracy} = \frac{\text{Recuperación} + \text{Especificidad}}{2}$$

3.5.2. *Precisión*

Esta métrica mide la proporción de predicciones positivas correctas entre todas las predicciones positivas realizadas (Berrar, 2019).

$$\text{Precisión} = \frac{VP}{VP + FP}$$

3.5.3. Recuperación

Esta métrica también se la conoce como sensibilidad y evalúa la proporción de casos positivos correctamente identificados entre todos los casos positivos reales (Berrar, 2019).

$$\text{Recuperación} = \frac{VP}{VP + FN}$$

3.5.4. Especificidad

Es una métrica que indica la proporción de casos negativos correctamente identificados entre todos los casos negativos reales (Berrar, 2019).

$$\text{Especificidad} = \frac{VN}{VN + FP}$$

3.5.5. F1-Score

Es la media armónica entre precisión y recuperación (Berrar, 2019). Es una métrica equilibrada que es útil cuando hay un desbalance entre las clases positivas y negativas.

$$F1 - Score = 2 \cdot \frac{\text{Precisión} \cdot \text{Recuperación}}{\text{Precisión} + \text{Recuperación}}$$

3.5.6. Área bajo la curva ROC (AUC-ROC)

Esta métrica resume el desempeño del modelo al medir su capacidad de distinguir entre clases positivas y negativas en diferentes umbrales de probabilidad.

3.6. Análisis de estabilidad en la selección de predictores

Con la finalidad de evaluar la robustez de los predictores seleccionados por los métodos de regularización (LASSO y Group LASSO), se optó por implementar un procedimiento llamado selección de estabilidad (en inglés, stability selection). Esta técnica permite estimar la probabilidad de que cada predictor sea retenido por el modelo cuando se perturba la muestra de entrenamiento (Meinshausen & Bühlmann, 2010).

Para este procedimiento, en primer lugar, se generaron $B = 100$ submuestras aleatorias a partir de la base de entrenamiento. Cada submuestra tuvo una fracción $\phi = 0.5$ de las observaciones originales. El muestreo fue sin reemplazo y en cada submuestra se ajustó con el modelo correspondiente según las estrategias definidas (LASSO o Group LASSO), con un mismo valor de penalización λ , que se obtuvo previamente mediante la validación

cruzada del conjunto completo de entrenamiento. Luego, se registraron los predictores seleccionados para cada ajuste; y, a partir de las repeticiones, se calculó para cada predictor la proporción de veces que fue seleccionado, es decir:

$$\hat{\pi}_j = \frac{1}{B} \sum_{b=1}^B 1 \{j \in S^{(b)}\}$$

Donde $S^{(b)}$ es el conjunto de variables seleccionadas en la repetición b .

Finalmente, se definió como una variable estable aquella que aparece como mínimo el 70% de las veces ($\pi_u = 0.7$), es decir, que cumple la condición $\hat{\pi}_j \geq \pi_u$.

La decisión de mantener fijo el parámetro de regularización se debe a dos razones: (i) reducir la carga computacional, al evitar recalcular la validación cruzada en cada submuestra, y (ii) aislar la fuente de variabilidad asociada únicamente al re-muestreo de observaciones.

3.7. Importancia Relativa Ponderada

Con la intención de evaluar la robustez de la relevancia de cada predictor cuando cambia el algoritmo de aprendizaje y la configuración de las estrategias se procede a calcular la importancia relativa ponderada de cada variable a través de los modelos. Es decir, para este análisis se opta por promediar la importancia relativa de cada variable (a través de modelos) y ponderar dicho promedio por el desempeño predictivo de cada modelo.

Para llevar a cabo esta evaluación, en primer lugar, para cada modelo $m = 1, \dots, M$, se obtiene un vector de importancia $I_m = (I_{m1}, I_{m2}, \dots, I_{mJ})$, de cada variable $j = 1, \dots, J$. Posteriormente, dado que cada algoritmo mide la importancia en escalas distintas, se procedió a normalizar dentro de cada modelo para que se encuentre en el rango $[0,100]$:

$$I_{mj}^* = 100 * \frac{I_{mj} - \min(I_m)}{\max(I_m) - \min(I_m)}$$

En el caso de que alguna variable no tenga importancia en el modelo, se registra con el valor de 0. De esta forma, se logra penalizar la ausencia y permite que el promedio refleje la importancia efectiva.

Adicionalmente, se construyen pesos. Para ello, se reúne las dos principales métricas de desempeño por modelo que son Balanced Accuracy y AUC, ambas normalizadas:

$$BAAC_m^{norm} = \frac{BAAC_m - \min(BAAC)}{\max(BAAC) - \min(BAAC)}$$

$$AUC_m^{norm} = \frac{AUC_m - \min(AUC)}{\max(AUC) - \min(AUC)}$$

Posteriormente, se construye el peso por modelo de una forma convexa:

$$w_m = 0.7 \cdot BAAC_m^{norm} + 0.3 \cdot AUC_m^{norm}$$

Se opta por esta construcción, dado que en un contexto de clases desbalanceadas, se recomienda el balanced accuracy porque combina la sensibilidad y la especificidad (Brodersen et al., 2010). Por su parte, el AUC es ampliamente aceptado como medida de calidad de los clasificadores, pero tiene limitaciones cuando el foco está en una clase minoritaria o en la calibración (Fawcett, 2006; Hand, 2009). Por lo tanto, se utiliza el AUC como medida complementaria, por eso tiene un menor peso.

Luego, se normaliza el peso de tal forma que $\sum_m w_m = 1$. Seguidamente, este peso se multiplica por el vector de importancia del modelo $w_m \cdot I_m$.

Finalmente, se calcula por cada variable j el promedio ponderado:

$$r_j = \sum_{m=1}^M w_m \cdot I_{mj}$$

El procedimiento implementado se ubica en un punto intermedio entre enfoques model-specific y model-agnostic. Por un lado, en modelos lineales (como la regresión logística y la SVM lineal), la importancia se deriva de la magnitud de los coeficientes estimados; en modelos basados en árboles (árbol de decisión, Random Forest y GBM), corresponde a la reducción de impureza o ganancia acumulada atribuible a cada predictor; mientras que en algoritmos sin una métrica nativa de importancia (KNN y SVM radial), se aproxima mediante medidas de sensibilidad en la precisión de clasificación. La estandarización realizada mediante `caret::varImp` permite trasladar todas estas medidas a una escala común, lo que facilita su integración en un índice de consenso. Esta lógica se encuentra estrechamente vinculada con los desarrollos más recientes de la literatura, en particular con el enfoque WISCA (Weighted Scaled Consensus Attributions) propuesto por Banegas-Luna et al. (2025), que formaliza la construcción de interpretaciones consensuadas a partir de modelos heterogéneos.

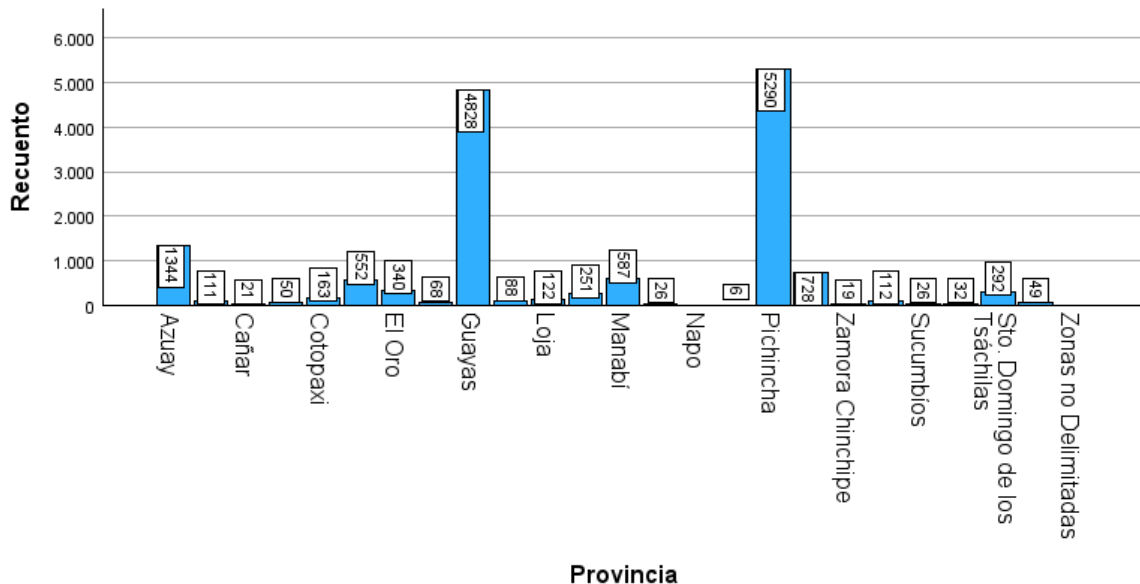
4. RESULTADOS

4.1. Estadísticas descriptivas

En primer lugar, antes de aplicar los modelos de selección y predicción, se realizó un análisis exploratorio de la base con el fin de caracterizar la distribución de las empresas y su comportamiento en la inversión en TIC. De esta manera, se permite dimensionar la heterogeneidad de la muestra. Cabe enfatizar que, para esta sección, se utilizan las estadísticas descriptivas ponderadas para asegurar representatividad poblacional.

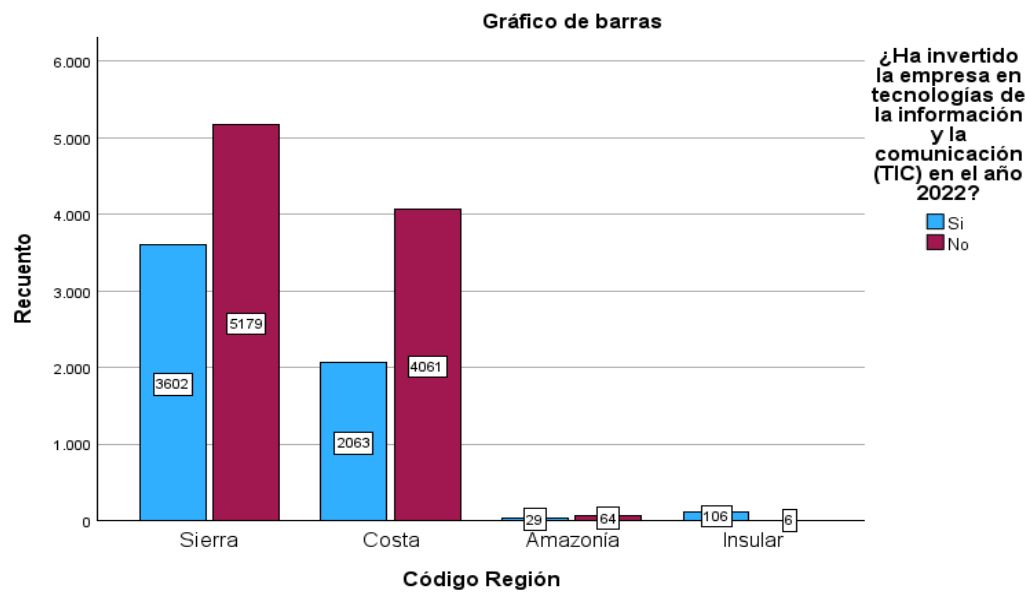
En cuanto a la distribución geográfica, en la Ilustración 1 se observa una alta concentración empresarial en las provincias de Pichincha (35 %) y Guayas (32%), que en conjunto agrupan más del 67% de las empresas. El resto de las provincias tiene una participación muy reducida, inferior al 9% cada una, lo que evidencia una marcada centralización territorial de la actividad económica.

Ilustración 1 - Número de empresas por provincia.



Siguiendo con la descripción geográfica, en la Ilustración 2 se observa que la región de la sierra abarca el 58.1% de la muestra, seguida por el 40.5% de la región costa, luego la región amazónica con el 0.6% y finalmente la región insular con apenas el 0.7%. En cuanto a la decisión de adoptar las TIC, curiosamente se observa que el 94.6% de las empresas de la región insular han decidido invertir en TIC. Mientras tanto, en la sierra, costa y amazonía predomina la no inversión, aunque con intensidades diferentes.

Ilustración 2 - Adopción de TIC por región.



Para analizar más de cerca estas diferencias en la variable región, se optó por realizar la prueba de chi-cuadrado de independencia. Esta prueba evalúa si la distribución de frecuencias observada en una tabla de contingencia difiere de la esperada bajo la hipótesis nula de independencia entre variables (Mangeaud et al., 2024). Como resultado, en la Tabla 4 se confirma que las diferencias observadas en la adopción de TIC son estadísticamente significativas ($\chi^2(3) = 234.85$; $p < 0.001$). Esto indica que la localización geográfica de las empresas se asocia de manera sistemática con su probabilidad de invertir en tecnologías digitales, aunque la magnitud y dirección de esta relación deben interpretarse a partir de las distribuciones específicas por región.

Tabla 4 - Resultados de la prueba de chi-cuadrado para la relación entre la región e inversión en TIC.

	Valor	gl	Significación asintótica (bilateral)
Chi-cuadrado de Pearson	234.85 ^a	3	<0.001
Razón de verosimilitud	247.83	3	<0.001
Asociación lineal por lineal	11.60	1	<0.001
N de casos válidos	15,110		

Nota: ^a0 casillas (0.0%) han esperado un recuento menor que 5. El recuento mínimo esperado es 35.70.

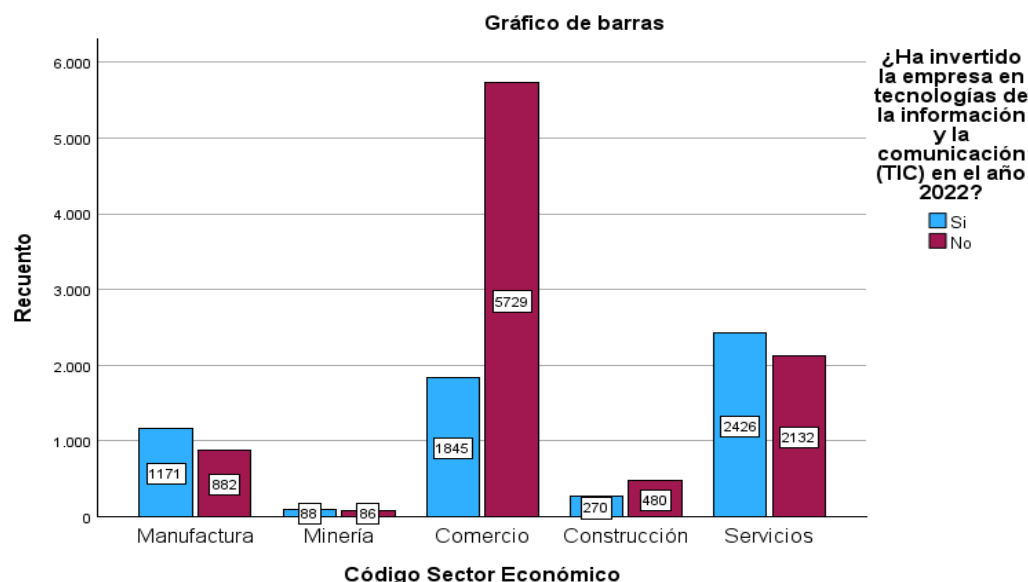
En la Tabla 5 se muestra la distribución de las firmas por tamaño y actividad económica. Con respecto al tamaño empresarial, según la clasificación del Tomo I, el 77% de la base corresponde a las empresas medianas tipo A (44%) y B (33%), mientras que, el 23% corresponde a las empresas grandes. Adicionalmente, al analizar la composición sectorial, se observa que el comercio (50%) y los servicios (30%) son los sectores predominantes, seguidos por la manufactura (14%), mientras que minería y construcción representan participaciones mucho menores, 1% y 5%, respectivamente.

Tabla 5 - Distribución de las firmas por tamaño y actividad económica.

Descripción	Mediana	Mediana	Grande	Total
	A	B		
Manufactura	774	564	715	2,053
Minería	63	38	74	175
Comercio	3,128	2,793	1,653	7,574
Construcción	381	254	115	750
Servicios	2,348	1,365	845	4,558
Total	6,694	5,014	3,402	15,110

Continuando con la estructura sectorial; en la Ilustración 3 se observa que, 5,800 (38.4%) firmas declararon haber invertido en TIC y 9,309 (61.6%) no realizaron inversión. El desglose sectorial revela que la manufactura presenta la mayor tasa de adopción de TIC (57.0%), seguida de los servicios (53.2%) y la minería (50.6%), lo cual refleja una mayor propensión relativa a invertir en tecnologías digitales en comparación con otros sectores. Esta información es relevante, dado que sugiere que la propensión a invertir en tecnologías digitales podría estar vinculada con la naturaleza productiva de cada sector.

Ilustración 3 - Adopción de TIC por sector económico.



Finalmente, para contrastar si las diferencias observadas en la adopción de TIC entre sectores económicos son estadísticamente significativas, se aplicó una prueba de chi-cuadrado de independencia (χ^2). En la Tabla 6 se observa que el estadístico obtenido ($\chi^2(4) = 1369.11$; $p < 0,001$) permite rechazar la hipótesis nula y concluir que la propensión a invertir en TIC varía significativamente según el sector económico. Esto confirma que la propensión a invertir en TIC está asociada de manera sistemática con la estructura sectorial, lo cual refuerza la relevancia de considerar esta variable en los modelos de predicción.

Tabla 6 - Resultados de la prueba de chi-cuadrado para la relación entre sector económico e inversión en TIC.

	Valor	gl	Significación asintótica (bilateral)
Chi-cuadrado de Pearson	1,369.11 ^a	4	<0.001
Razón de verosimilitud	1,386.74	4	<0.001
Asociación lineal por lineal	50.12	1	<0.001
N de casos válidos	15,109		

Nota: ^a0 casillas (0.0%) han esperado un recuento menor que 5. El recuento mínimo esperado es 66.79.

4.2. Predictores seleccionados

En la Tabla 7 se reporta el Error Cuadrático Medio (ECM) comparativo entre las distintas estrategias de selección de predictores. Los resultados muestran que, en la estrategia de Regularización sin categóricas multinivel, Lasso reduce el ECM respecto a Ridge (0.419 vs. 0.473), lo que evidencia la eficiencia de la penalización L_1 en contextos de alta dimensionalidad. Cuando se incorporan variables categóricas multinivel, se emplea Group Lasso que, si bien es el método metodológicamente correcto, no mejora sustancialmente a Ridge (ECM de 0.472 y 0.450 frente a 0.473). En la Estrategia Híbrida ocurre algo similar; con las categóricas multinivel, Group Lasso logra un ECM de 0.429 frente al 0.478 de Ridge, y sin categóricas multinivel, Lasso alcanza valores prácticamente iguales (0.429–0.430). En conjunto, los hallazgos muestran que Ridge mantiene un desempeño estable pero menos competitivo, mientras que Lasso y Group Lasso logran reducciones del ECM en la todas las configuraciones, con especial ventaja para Lasso cuando no intervienen variables categóricas.

Tabla 7 - Resultados del Error de Cuadrático Medio (Lasso vs. Group Lasso) por estrategia y número de particiones.

Estrategia	Variante	Particiones CV	ECM		
			Ridge	LASSO	GLASSO
Regularización	Sin categóricas	10	0.473	0.419	
	multinivel	5	0.473	0.419	
	Con categóricas	10	0.473		0.472
	multinivel	5	0.473		0.450
Híbrida	Con categóricas	10	0.478		0.429
	multinivel	5	0.478		0.429
	Sin categóricas	10	0.478	0.430	
	multinivel	5	0.478	0.429	

En la Tabla 8 se resume el número de predictores seleccionados en cada estrategia y variante, considerando tanto la validación cruzada con 10 particiones como con 5 particiones. En el cuerpo principal se destacan únicamente el número total de predictores y

algunos ejemplos representativos de los más relevantes. Los resultados muestran que la estrategia de regularización automática retuvo un mayor número de variables, mientras que la estrategia de selección híbrida redujo significativamente el conjunto, alcanzando mayor parsimonia. En cambio, la estrategia de selección manual se mantuvo estable en ambos esquemas de validación, dado que los predictores fueron definidos a priori según criterios teóricos.

Tabla 8 - Resumen de predictores seleccionados por estrategia y tipo de validación cruzada.

Estrategia	Variante	Particiones CV	Cantidad de Predictores
Regularización	Sin categóricas multinivel	10	32
		5	32
	Con categóricas multinivel	10	88
		5	77
Selección Manual	Sin categóricas multinivel	10	24
		5	24
	Con categóricas multinivel	10	17
		5	17
Híbrida	Sin categóricas multinivel	10	17
		5	16
	Sin categóricas multinivel	10	10
		5	11

Del mismo modo, destacamos que las listas completas de los predictores seleccionados por cada estrategia se incluyen en el Anexo A. Tablas de predictores. Al comparar las tres estrategias, se observa que ciertos predictores aparecen de forma recurrente en la mayoría de las configuraciones, lo que refuerza su papel central en la explicación de la inversión en TIC. En la Tabla 9, se presentan los principales 10 predictores que aparecen en las configuraciones iniciales. Entre ellos, se destacan la variable de innovación, la disponibilidad de un sitio web, la edad de la firma y la presencia de personal especialista en

TIC, que fueron retenidos en todas las estrategias y esquemas de validación. Otros predictores, como las actividades de compras en línea y el nivel de endeudamiento del patrimonio, aparecen de manera más recurrente en las estrategias manuales o híbridas. En general, más allá de las variaciones metodológicas entre estrategias y particiones, existe un conjunto reducido de predictores que emerge de manera sistemática como explicativo de las decisiones de inversión en TIC por parte de las firmas.

Tabla 9 - Predictores con mayor frecuencia de selección en las doce configuraciones.

No.	Predictores	Recuento	Descripción
1	Innovación	12	Indica si la empresa realizó actividades de investigación y desarrollo orientados a la generación de productos/procesos nuevos o significativamente mejorados.
2	PA_Especialista_TIC	12	Indica si tiene personal afiliado especialista TIC.
3	Sitio_Web	12	Indica si la empresa disponía de sitio o página web.
4	Edad	12	Indica la edad de la empresa
5	Compras_por_Internet	10	Indica si la empresa compró bienes o servicios por internet.
6	Endeudamiento_Patrimonio	9	Indica el ratio del nivel de endeudamiento del patrimonio.
7	VAE	8	Indica el valor agregado empresarial.
8	Total_PA_Mujeres	8	Indica el total de mujeres.
9	ID_Externa	7	Indica el valor que destinó la empresa a Investigación y Desarrollo (I+D) externa.
10	FirmaDigital	7	Indica si la empresa usó firma digital en comunicaciones enviadas.

4.3. Desempeño predictivo

Una vez identificados los conjuntos de predictores retenidos en cada estrategia, el siguiente paso consiste en evaluar su capacidad para predecir la inversión en TIC. Para ello, se aplicaron los siete algoritmos de aprendizaje automático bajo cada configuración de variables, comparando su desempeño a partir de las métricas de clasificación. Este análisis permite observar la estrategia de selección que resulta más efectiva en términos predictivos, así como, examinar la estabilidad de los resultados bajo diferentes esquemas de validación cruzada.

Los principales resultados del desempeño predictivo de los modelos se presentan en la Tabla 10, mientras que el detalle completo de las métricas por cada configuración se encuentra en el apartado de Anexo B. Tablas de desempeño predictivo. Como se puede observar, ninguna estrategia domina de manera absoluta en todos los escenarios, pero se identifican patrones sólidos. El mayor valor de BAAC (0.759) se obtuvo con el modelo SVM lineal en la estrategia de regularización con categóricas (Group LASSO, CV=5). En términos de AUC, el mejor resultado (0.814) correspondió a RF bajo la estrategia de regularización sin categóricas (LASSO, CV=10). En general, los algoritmos que alcanzaron con mayor frecuencia el mejor desempeño fueron RF y SVM lineal (cuatro configuraciones cada uno), mientras que la regresión logística solo se destacó dentro de la estrategia de selección manual.

Los valores de BAAC de los mejores modelos se encuentran en un rango acotado, entre 0.703 y 0.759, sugiere que, independientemente de la estrategia de selección de predictores, la capacidad de los modelos para clasificar correctamente empresas que invierten y no invierten en TIC (ajustando por desbalance) fue equilibrado. Mientras tanto, el AUC mostró mayor dispersión, con valores entre 0.723 y 0.814, lo que sugiere que la capacidad de discriminación entre ambos grupos depende más del algoritmo y de la estrategia de selección de predictores. En conjunto, los resultados sugieren que la estrategia de selección manual proporciona un rendimiento robusto y estable con la ventaja de parsimonia, mientras que la regularización automática maximiza la precisión predictiva a costa de complejidad.

Estos resultados sugieren que, aunque los métodos de ensamble (RF) y los modelos lineales regulares (SVM) se alternan en el liderazgo del desempeño, sus diferencias en BAAC son relativamente estrechas. Ello indica que la robustez predictiva no depende exclusivamente del algoritmo, sino también de la estrategia de selección de predictores

aplicada. Por otra parte, el hecho de que la regresión logística solo destaque en la selección manual resalta la importancia de la fundamentación teórica en la elección de predictores, ya que permite a un modelo simple acercarse al rendimiento de métodos más sofisticados.

Tabla 10 - Desempeño predictivo por estrategia, variante y tamaño de partición para validación cruzada.

Estrategia	Variante	Particiones CV	Cantidad de Predictores	Mejor modelo	BAAC	AUC
Regularización	Sin categóricas (LASSO)	10	32	Random Forest	0.751	0.814
		5	32	Random Forest	0.745	0.804
	Con categóricas (Group LASSO)	10	88	SVM lineal	0.745	0.745
		5	77	SVM lineal	0.759	0.759
	Con categóricas	10	24	Logístico	0.738	0.788
		5	24	Logístico	0.738	0.788
	Sin categóricas	10	17	Logístico	0.736	0.792
		5	17	Logístico	0.736	0.792
Selección Manual						
Híbrida	Con categóricas (Group LASSO)	10	17	SVM lineal	0.723	0.723
		5	16	Random Forest	0.729	0.783
	Sin categóricas (LASSO)	10	10	Random Forest	0.703	0.769
		5	11	Random Forest	0.730	0.783

4.4. Análisis de Estabilidad en la selección de predictores

La salida del procedimiento de selección de estabilidad son las estimaciones $\hat{\pi}_j$, que cuantifican la probabilidad de selección de cada predictor bajo perturbaciones en la muestra. Estas medidas se utilizaron para identificar el subconjunto de predictores más estables dentro de cada estrategia de regularización, complementando así el análisis de desempeño predictivo de los modelos. Los resultados completos, se pueden encontrar en Anexo C. Análisis de Estabilidad.

A continuación, en la Tabla 11 se sintetiza los 10 predictores con mayor estabilidad en las 8 configuraciones de las dos estrategias que involucran la regularización. Los resultados muestran un núcleo de variables altamente sólidos: la innovación y la cantidad de especialistas TIC con los que cuenta la empresa (PA_Especialista_TIC) que alcanzan valores cercanos a 1 en todas las variantes (promedio 0.995), así como, el hecho de disponer un sitio web (0.914) lo que indica que son prácticamente invariantes al tipo de partición o al tratamiento de categóricas. En un segundo nivel se ubican las variables de Compras_por_Internet (0.680) y Edad (0.633) que, si bien no alcanzan la estabilidad absoluta, muestran una recurrencia sistemática en las distintas variantes. Mientras tanto, existen variables como Año_compra_afi y Año_compra_software, Red_LAN y Uso_Internet_Capacitación que muestran estabilidad intermedia, lo que sugiere que su inclusión depende en mayor medida de la estrategia de selección aplicada.

Tabla 11 - Top 10 de predictores estables en las estrategias de regularización e híbrida.

No.	Variables	Regularización				Híbrida				Promedio
		SC	SC	CC	CC	CC	CC	SC	SC	
5	10	5	10	5	10	5	10	5	10	
1	Innovación	0.98	0.98	1	1	1	1	1	1	0.995
2	PA_Especialista_TIC	0.98	0.98	1	1	1	1	1	1	0.995
3	Sitio_Web	0.81	0.81	0.84	0.85	1	1	1	1	0.914
4	Compras_por_Internet	0.72	0.72			1	1	1	1	0.680
5	Edad			0.84	0.83	0.97	0.97	0.74	0.71	0.633

6	Año_compra _afi	0.98	0.98	1	1	0.495
7	Año_compra _software	0.98	0.98	1	1	0.495
8	Red_LAN Uso_Internet	0.98	0.98	1	1	0.495
9	_Capacitacio n	0.97	0.97	1	1	0.493
10	Compras_por _Internet_por centaje	0.94	0.94	1	1	0.485

Estos resultados revelan un núcleo robusto de predictores: la innovación, especialistas TIC y sitio web; cuya presencia constante sugiere que representan determinantes estructurales de la inversión en TIC, más allá de la estrategia metodológica aplicada. Esta diferenciación entre predictores altamente estables y otros más sensibles a la configuración metodológica permite jerarquizar las variables, destacando aquellas que pueden considerarse determinantes centrales en la explicación de la inversión frente a aquellas cuyo papel es más contingente.

4.5. Importancia Relativa Ponderada de los predictores

Esta valoración se la realizó con el objetivo de identificar las variables que aportan mayor capacidad explicativa en la predicción de la inversión en TIC, más allá de las diferencias en desempeño entre modelos y estrategias. De esta forma, se puede observar la robustez de los predictores claves y se evalúa si los hallazgos empíricos coinciden con las expectativas teóricas derivadas de la literatura. Los resultados completos por cada estrategia y variante se incluyen en el Anexo D. Importancia Relativa Ponderada, donde se reportan: (i) las tablas de importancia ponderada de los predictores y (ii) un mapa de calor con la importancia ponderada de los 10 predictores principales en cada configuración.

La Tabla 12 muestra el ranking de los 15 predictores con mayor importancia relativa ponderada promedio a través de las doce configuraciones de modelos y estrategias. Este

criterio penaliza a las variables que solo fueron consideradas en estrategias específicas (manual o híbrida), no obstante, garantiza que la jerarquización refleje la relevancia global de los predictores.

Los resultados confirman que los factores más influyentes están relacionados con la capacidad innovadora de la empresa (*Innovación*), la *disponibilidad de personal TIC especializado* (*PA_Especialista_TIC*), y aspectos de infraestructura digital básica (*Sitio_Web*, *Certificado_Digital*, *Red_LAN*). A estos se suman variables vinculadas al capital humano femenino (*PA_ApoyoAd_Mujeres*, *Salarios de mujeres* en diferentes cortes), que aparecen con valores altos en varias configuraciones, lo que resalta la importancia del empleo y remuneración del personal como determinantes de la adopción tecnológica. Finalmente, algunas variables financieras y de trayectoria, como el endeudamiento del patrimonio, la edad de la empresa y el año de compra de software, completan el conjunto de predictores robustos, mostrando que la decisión de invertir en TIC responde a una combinación de factores tecnológicos, organizativos y financieros cuya relevancia se mantiene estable en diferentes modelos y escenarios.

Tabla 12 – Top 15 predictores con mayor importancia relativa ponderada promedio.

No.	Variables	Regularización				Manual				Híbrido				Promedio
		SC	SC	CC	CC	CC	CC	SC	SC	CC	CC	SC	SC	
1	Innovación	10.23	11.36	3.72	3.78	1.66	1.66	1.56	1.48	3.75	3.10	3.02	3.12	4.04
2	PA_Especialista_TIC	6.52	6.04	3.86	4.39	2.09	2.09	1.93	1.84	3.74	2.97	3.10	2.95	3.46
3	PA_ApoyoAd_Mujeres	13.46	14.10	5.07	5.17									3.15
4	Sitio_Web	5.01	5.01	3.04	3.41	1.87	1.87	1.61	1.53	3.13	2.64	2.43	2.62	2.85
5	Salarios_Nov_ApoyoAd_Mujeres	12.71	13.79	5.05	0.36									2.66
6	Salarios_Mujeres	14.99	14.93											2.49
7	Salarios_Nov_Mujeres	15.30	14.47											2.48
8	Año_compra_software	7.72	8.56	3.86	3.80									2.00
9	Total_PA_Mujeres					1.08	1.09	1.04	1.02	4.47	4.14	4.15	4.31	1.78
10	Endeudamiento_Patrimonio			1.09	5.37	1.14	0.77	0.89	0.81	4.09	3.00	2.97		1.68
11	Edad	3.04	2.96	1.92	2.04	1.00	1.00	0.70	0.67	1.86	1.55	1.32	1.35	1.62
12	Compras_por_Internet	3.09	3.62			1.15	1.16	1.00	0.96	2.00	1.79	1.78	1.85	1.53
13	Certificado_Digital	4.90	5.36	3.22	3.99									1.46
14	VAE					0.78	0.78	0.67	0.66	3.93	3.25	3.42	3.58	1.42
15	Total_Dispositivos									4.57	3.91	3.79	4.17	1.37

SC: Sin categóricas multinivel, CC: Con categóricas multinivel

5. DISCUSIÓN Y CONCLUSIONES

Esta investigación tuvo como objetivo identificar los factores que explican y permiten predecir la probabilidad de que las empresas ecuatorianas inviertan en Tecnologías de la Información y Comunicación (TIC) en 2022, utilizando información de la encuesta ENESEM y aplicando un marco metodológico que combina la estadística descriptiva, estrategias de selección de variables (regularización, manual e híbrida) y algoritmos de aprendizaje automático. La investigación se planteó como una contribución tanto al análisis empírico de la adopción digital en el ámbito empresarial como a la discusión metodológica. Cabe destacar que, un aspecto central del diseño fue la evaluación de variantes metodológicas donde se compararon escenarios que incluyeron o excluyeron variables categóricas multinivel, así como configuraciones con diferentes particiones de validación cruzada (10 y 5). Este enfoque permitió no solo contrastar el desempeño predictivo bajo distintos supuestos, sino también analizar la robustez y estabilidad de los predictores seleccionados frente a cambios en la estrategia de modelado.

Con respecto a nuestros resultados, en primer lugar, las estadísticas descriptivas mostraron la heterogeneidad estructural de la muestra en términos de tamaño, sector y localización geográfica. La concentración empresarial en provincias como Pichincha y Guayas confirma la centralidad histórica de estos polos económicos en la actividad productiva nacional, en línea con lo señalado por la literatura sobre aglomeración empresarial (Tobar-Pesántez & Solano-Gallegos, 2017). Del mismo modo, aunque el comercio concentra el mayor número de empresas, son los sectores de servicios (53.2% de sus firmas), manufactura (57.0%) y minería (50.6%) los que muestran una mayor propensión relativa a invertir en TIC. Esta heterogeneidad sectorial implica que los resultados deben interpretarse considerando tanto la composición de la muestra como las diferencias internas en patrones de adopción. En términos prácticos, este hallazgo sugiere que las políticas de fomento digital no deberían ser uniformes, por ejemplo, en los sectores de servicios y manufactura resulta más pertinente profundizar en la sofisticación de usos; mientras que, en sectores rezagados como la construcción, en cambio, la prioridad debería ser garantizar la inclusión digital básica para cerrar brechas estructurales.

Por otra parte, en cuanto a la selección de predictores, se compararon las tres estrategias. A pesar de las diferencias en la cantidad de variables retenidas (desde 17 hasta 88), existe un núcleo sólido de predictores que aparecen de manera recurrente: Innovación,

disponibilidad de especialistas TIC, existencia de sitio web, uso de internet para capacitación, y porcentaje de compras por internet. Este hallazgo coincide con la literatura en economía de la innovación, que resalta la capacidad tecnológica y los recursos humanos especializados como factores clave en la adopción digital de las empresas en países en desarrollo (Arvanitis et al., 2013; Bresnahan et al., 2002). Desde una perspectiva práctica, esta jerarquía de predictores implica que las políticas públicas deberían priorizar el fortalecimiento de las capacidades de innovación y la formación de talento especializado, mientras que a nivel empresarial se subraya la importancia de mantener una presencia digital básica (sitio web) como condición mínima para avanzar hacia procesos de transformación digital.

En cuanto a la estrategia de selección manual se observó que los resultados permanecieron inalterados al modificar el número de particiones en la validación cruzada (5 y 10). Este hallazgo aporta un matiz relevante; mientras la regularización posibilita identificar patrones adicionales no previstos inicialmente, la selección manual garantiza estabilidad, aunque limitada al marco teórico original. Desde un punto de vista metodológico, el contraste entre ambas estrategias constituye una contribución del estudio, al mostrar cómo la elección del mecanismo de selección condiciona la sensibilidad de los resultados. En términos prácticos, la congruencia de la estrategia manual puede resultar útil para aplicaciones donde se prioriza la replicabilidad y la transparencia del criterio de selección. No obstante, su limitación al no incorporar factores emergentes implica que para fines de política pública o gestión empresarial es recomendable complementarla con métodos flexibles, capaces de captar cambios en los determinantes de la inversión en TIC.

Adicionalmente, en relación con el desempeño predictivo, los modelos basados en árboles de decisión ensamblados (Random Forest) alcanzaron los niveles más altos de precisión balanceada y AUC, superando de forma persistente a los modelos lineales. Este resultado coincide con la literatura que resalta la capacidad de los algoritmos no lineales para capturar estructuras complejas en contextos de alta dimensionalidad (Wu & Bajwa, 2015; Yamada et al., 2018). No obstante, el buen desempeño observado en el SVM lineal sugiere que también existe un componente lineal relevante en la relación entre los predictores y la inversión en TIC. En conjunto, estos hallazgos muestran que dicha relación no es estrictamente lineal, sino que incorpora efectos combinados y dependencias cruzadas que requieren de algoritmos flexibles para su adecuada modelización. Desde una perspectiva práctica, esto implica que tanto investigadores como responsables de política deberían optar

por enfoques que integren componentes lineales y no lineales, a fin de identificar patrones de complementariedad entre características empresariales al diseñar estrategias de fomento a la inversión en TIC.

El análisis de selección de estabilidad reforzó la robustez de predictores como Innovación y Especialistas TIC, ya que aparecieron con alta frecuencia de selección en diferentes configuraciones de muestreo y validación cruzada. Por su parte, el análisis de importancia relativa ponderada mostró que estos predictores mantienen una posición destacada dentro del top 15, lo que confirma su relevancia de manera comparativa entre modelos y estrategias. Este resultado coincide con estudios previos que destacan la estabilidad de la innovación y del capital humano especializado como determinantes de la adopción tecnológica (Arvanitis et al., 2011; Bresnahan et al., 2002).

La presencia recurrente de estos factores (innovación y especialistas TIC) aporta un hallazgo adicional: en lugar de comportarse como sustitutos, actúan como dimensiones complementarias que reflejan la capacidad productiva de la firma para generar y absorber conocimiento. Empíricamente, esta doble evidencia; derivada de la selección por estabilidad y de la importancia relativa ponderada; refuerza la coherencia de los resultados más allá de la estrategia de selección utilizada. Metodológicamente, el contraste entre ambos enfoques constituye una contribución del estudio, al mostrar cómo diferentes criterios de evaluación convergen hacia un mismo conjunto de predictores claves. Desde el plano práctico, la inversión en TIC debe ser concebida no como un gasto aislado en equipos o software, sino como parte de una estrategia de transformación organizacional que combine capital humano especializado e innovación, elementos centrales para sostener la competitividad empresarial en el largo plazo.

Asimismo, resulta notable la aparición de un conjunto de variables vinculadas al género, particularmente aquellas relacionadas con la remuneración y la presencia femenina en áreas específicas de la empresa, lo cual sugiere que la participación laboral femenina y su remuneración pueden estar asociadas con las decisiones de inversión en TIC. Este resultado conecta con estudios previos que han reportado efectos heterogéneos de la diversidad de género sobre la innovación y la adopción tecnológica (Anggraini et al., 2023; Orser & Riding, 2018; Živković et al., 2024), pero aporta evidencia novedosa al mostrar que, en el contexto ecuatoriano, tales variables emergen como predictores relevantes. Desde el punto de vista metodológico, este hallazgo es destacable porque surge de un análisis diseñado para identificar predictores robustos, lo que sugiere que los patrones de género no son un artefacto

estadístico, sino un posible determinante estructural de la inversión digital. En términos prácticos, esta evidencia abre una línea de investigación y acción de política pública orientada a comprender cómo la composición de la fuerza laboral y las brechas de género influyen en la adopción tecnológica.

6. LIMITACIONES Y RECOMENDACIONES

Esta investigación presenta limitaciones que deben considerarse al interpretar los resultados. En primer lugar, el análisis se concentró en un único año de referencia (2022), lo que impide identificar dinámicas temporales o cambios en los determinantes de la inversión en TIC. Futuras investigaciones podrían ampliar el horizonte temporal con series de varios años y en distintos contextos sectoriales, con el fin de evaluar la estabilidad de los predictores y capturar patrones de inversión más completos.

En segundo lugar, aunque se compararon variantes metodológicas con y sin variables categóricas multinivel y con validación cruzada de 10 y 5 pliegues, la selección de predictores depende de la información disponible en la encuesta. Esto implica que la ausencia de variables categóricas multiniveles relevantes podría influir en el rendimiento comparativo de los algoritmos. Además, ciertos factores no observados, como la capacidad gerencial, la cultura organizacional o las actitudes hacia la adopción tecnológica, pueden haber quedado fuera del análisis. Incluir estas dimensiones en futuros estudios permitiría enriquecer la explicación de los determinantes de la inversión digital.

En tercer lugar, el enfoque adoptado se centra en la predicción y, en consecuencia, se limita a identificar asociaciones estadísticas entre las variables, sin estimar la magnitud ni la dirección de los efectos, ni analizar las interacciones o establecer relaciones causales. Dado el carácter transversal de la información y la ausencia de variación exógena, no fue posible establecer inferencias causales en este estudio. No obstante, futuras investigaciones podrían avanzar en esta dirección mediante el uso de datos longitudinales, técnicas de inferencia causal o herramientas para la interpretación de los modelos, lo que permitiría una comprensión más profunda de los factores que condicionan la inversión en TIC.

Finalmente, dado que este estudio se centró en Ecuador, una línea futura de investigación consiste en contrastar los resultados con evidencia de otros países en desarrollo o de la región latinoamericana. Esto permitiría evaluar si los patrones observados, incluida la relevancia de factores de género en la inversión en TIC, son específicos del contexto ecuatoriano o forman parte de tendencias más amplias.

REFERENCIAS

- Achanta, M. (2024). The Impact of Real - Time Data Processing on Business Decision - making. *International Journal of Science and Research (IJSR)*, 13(7), 400–404. <https://doi.org/10.21275/SR24708033511>
- Almerich, G., Suárez-Rodríguez, J., Díaz-García, I., & Orellana, N. (2021). The influence of using ICT in high-skills competences and ICT competences. A structural model. *Education and Information Technologies*, 26(4), 3845–3869. <https://doi.org/10.1007/s10639-021-10437-y>
- Alvarado Vazquez, S., Madureira, A. M., Ostermann, F. O., & Pfeffer, K. (2023). The Use of ICTs to Support Social Participation in the Planning, Design and Maintenance of Public Spaces in Latin America. *ISPRS International Journal of Geo-Information*, 12(6), 237. <https://doi.org/10.3390/ijgi12060237>
- Anggraini, F., Putri, D., Novianti, N., & Helmi, S. (2023). Effect of Women's Entrepreneurship Practices on Information Communication Technology (ICT) Adoption During the Covid-19 Pandemic: A Case Study in West Sumatra, Indonesia. *Journal of Economics, Finance And Management Studies*, 06(06). <https://doi.org/10.47191/jefms/v6-i6-65>
- Arvanitis, S., Loukis, E., & Diamantopoulou, V. (2011). The impact of different types of ICT on innovation performance of Greek firms. *European, Mediterranean & Middle Eastern Conference on Information Systems*, 30–31.
- Arvanitis, S., Loukis, E., & Diamantopoulou, V. (2013). The effect of soft ICT capital on innovation performance of Greek firms. *Journal of Enterprise Information Management*, 26(6), 679–701. <https://doi.org/10.1108/JEIM-07-2013-0048>
- Banegas-Luna, A. J., Pérez-Sánchez, H., & Martínez-Cortés, C. (2025). *WISCA: A Consensus-Based Approach to Harmonizing Interpretability in Tabular Datasets*. <https://doi.org/https://doi.org/10.48550/arXiv.2506.06455>
- Belfo, F., Trigo, A., & Estébanez, R. P. (2015). Impact of ICT Innovative Momentum on Real-Time Accounting. *Business Systems Research Journal*, 6(2), 1–17. <https://doi.org/10.1515/bsrj-2015-0007>
- Ben Youssef, A., Hadhri, W., & M'Henni, H. (2011). Intra-Firm Diffusion of Innovation: Evidence from Tunisian SMEs Regarding Information and Communication Technologies. *Middle East Development Journal*, 3(1), 75–97.

- <https://doi.org/10.1142/S1793812011000338>
- Berrar, D. (2019). Performance Measures for Binary Classification. In *Encyclopedia of Bioinformatics and Computational Biology* (pp. 546–560). Elsevier. <https://doi.org/10.1016/B978-0-12-809633-8.20351-8>
- Bolliger, D. U., & Halupa, C. (2018). Online student perceptions of engagement, transactional distance, and outcomes. *Distance Education*, 39(3), 299–316. <https://doi.org/10.1080/01587919.2018.1476845>
- Breiman, L. (2001). Random Forests. *Machine Learning*, 45(1), 5–32. <https://doi.org/10.1023/A:1010933404324>
- Bresnahan, T. F., Brynjolfsson, E., & Hitt, L. M. (2002). Information Technology, Workplace Organization, and the Demand for Skilled Labor: Firm-Level Evidence. *The Quarterly Journal of Economics*, 117(1), 339–376. <https://doi.org/10.1162/003355302753399526>
- Brodersen, K. H., Ong, C. S., Stephan, K. E., & Buhmann, J. M. (2010). The Balanced Accuracy and Its Posterior Distribution. *2010 20th International Conference on Pattern Recognition*, 3121–3124. <https://doi.org/10.1109/ICPR.2010.764>
- Cardona, M., Kretschmer, T., & Strobel, T. (2013). ICT and productivity: conclusions from the empirical literature. *Information Economics and Policy*, 25(3), 109–125. <https://doi.org/10.1016/j.infoecopol.2012.12.002>
- Chang, V., Xu, Q. A., Akinloye, S. H., Benson, V., & Hall, K. (2024). Prediction of bank credit worthiness through credit risk analysis: an explainable machine learning study. *Annals of Operations Research*. <https://doi.org/10.1007/s10479-024-06134-x>
- Chepurna, M., & Rialp Criado, J. (2018). Identification of barriers to co-create on-line: the perspectives of customers and companies. *Journal of Research in Interactive Marketing*, 12(4), 452–471. <https://doi.org/10.1108/JRIM-01-2018-0018>
- Cox, D. R. (1958). The Regression Analysis of Binary Sequences. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 20(2), 215–232. <https://doi.org/10.1111/j.2517-6161.1958.tb00292.x>
- Cuevas-Vargas, H., Enríquez, L. A., Adame, M. G., & Servin, J. L. (2015). The use of ICTs and its relation with the competitiveness of Mexican SMEs. *European Scientific Journal*, 11(13).
- Czajkowski, M., & Kretowski, M. (2019). Decision tree underfitting in mining of gene expression data. An evolutionary multi-test tree approach. *Expert Systems with*

- Applications*, 137, 392–404. <https://doi.org/10.1016/j.eswa.2019.07.019>
- Danquah, M., & Amankwah-Amoah, J. (2017). Assessing the relationships between human capital, innovation and technology adoption: Evidence from sub-Saharan Africa. *Technological Forecasting and Social Change*, 122, 24–33. <https://doi.org/10.1016/j.techfore.2017.04.021>
- Dave, B., & Koskela, L. (2009). Collaborative knowledge management—A construction case study. *Automation in Construction*, 18(7), 894–902. <https://doi.org/10.1016/j.autcon.2009.03.015>
- Demir, S., & Şahin, E. K. (2022). Evaluation of Oversampling Methods (OVER, SMOTE, and ROSE) in Classifying Soil Liquefaction Dataset based on SVM, RF, and Naïve Bayes. *European Journal of Science and Technology*. <https://doi.org/10.31590/ejosat.1077867>
- Díaz-Chao, Á., Sainz-González, J., & Torrent-Sellens, J. (2015). ICT, innovation, and firm productivity: New evidence from small local firms. *Journal of Business Research*, 68(7), 1439–1444. <https://doi.org/10.1016/j.jbusres.2015.01.030>
- Duc, D. T. V., & Nguyen, P. Van. (2023). ICT impact and firm size: Empirical results from Vietnam. *African Journal of Science, Technology, Innovation and Development*, 15(3), 337–348. <https://doi.org/10.1080/20421338.2022.2091421>
- Eom, T., Woo, C., & Chun, D. (2024). Predicting an ICT business process innovation as a digital transformation with machine learning techniques. *Technology Analysis & Strategic Management*, 36(9), 2271–2283. <https://doi.org/10.1080/09537325.2022.2132927>
- Eze, S. C., & Chinedu-Eze, V. C. (2018). Strategic roles of actors in emerging information communication technology (EICT) adoption in SMEs. *The Bottom Line*, 31(2), 114–136. <https://doi.org/10.1108/BL-09-2017-0029>
- Eze, S. C., & Vera, C. C.-E. (2018). Examining information and communication technology (ICT) adoption in SMEs. *Journal of Enterprise Information Management*, 31(2), 338–356. <https://doi.org/10.1108/JEIM-12-2014-0125>
- Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters*, 27(8), 861–874. <https://doi.org/10.1016/j.patrec.2005.10.010>
- Finkelstein Shapiro, A., Nuguer, V., & Novoa Gomez, S. (2024). Labor Market and Macroeconomic Dynamics in Latin America amid COVID: The Role of Digital-Adoption Policies. *The World Bank Economic Review*, 38(1), 161–184.

- <https://doi.org/10.1093/wber/lhad019>
- Fonti, V., & Belitser, E. (2017). Feature selection using lasso. *VU Amsterdam Research Paper in Business Analytics*, 30, 1–25.
- Friedman, J. (2013). *Tools of the trade: when to use those sample weights*. World Bank Blogs. <https://blogs.worldbank.org/en/impactevaluations/tools-of-the-trade-when-to-use-those-sample-weights>
- Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. *Annals of Statistics*, 1189–1232. <http://www.jstor.org/stable/2699986>.
- García-Álvarez, M. T. (2015). Analysis of the effects of ICTs in knowledge management and innovation: The case of Zara Group. *Computers in Human Behavior*, 51, 994–1002. <https://doi.org/10.1016/j.chb.2014.10.007>
- Gheisari, G. (2024). The role of information and communication technology (ICT) in facilitating and accelerating international business processes. *International Journal of Research Publication and Reviews*, 4319–4326. <https://doi.org/10.55248/gengpi.5.0624.1550>
- Ghosh, S., Dasgupta, A., & Swetapadma, A. (2019). A Study on Support Vector Machine based Linear and Non-Linear Pattern Classification. *2019 International Conference on Intelligent Sustainable Systems (ICISS)*, 24–28. <https://doi.org/10.1109/ISS1.2019.8908018>
- Gimenez, G., & Vargas-Montoya, L. (2021). ICT Use and Successful Learning: The Role of the Stock of Human Capital. *Mathematics*, 9(14), 1648. <https://doi.org/10.3390/math9141648>
- Giotopoulos, I., Kontolaimou, A., Korra, E., & Tsakanikas, A. (2017). What drives ICT adoption by SMEs? Evidence from a large-scale survey in Greece. *Journal of Business Research*, 81, 60–69. <https://doi.org/10.1016/j.jbusres.2017.08.007>
- Gomez-Sanchez, A. M., Rosero-Ceballos, E. L., & Mosquera, D. A. (2023). Imports-ICT Linkage in Colombian Manufacturing. *International Journal of the Economics of Business*, 30(2), 185–205. <https://doi.org/10.1080/13571516.2023.2168996>
- Grazzi, M., & Jung, J. (2019). What are the drivers of ICT diffusion? Evidence from Latin American firms. *Information Technologies & International Development*, 15, 15.
- Guerrieri, P., Luciani, M., & Meliciani, V. (2011). The determinants of investment in information and communication technologies. *Economics of Innovation and New Technology*, 20(4), 387–403.

- Hagsten, E., & Kotnik, P. (2017). ICT as facilitator of internationalisation in small- and medium-sized firms. *Small Business Economics*, 48(2), 431–446. <https://doi.org/10.1007/s11187-016-9781-2>
- Hall, B. H., Lotti, F., & Mairesse, J. (2013). Evidence on the impact of R&D and ICT investments on innovation and productivity in Italian firms. *Economics of Innovation and New Technology*, 22(3), 300–328. <https://doi.org/10.1080/10438599.2012.708134>
- Hand, D. J. (2009). Measuring classifier performance: a coherent alternative to the area under the ROC curve. *Machine Learning*, 77(1), 103–123. <https://doi.org/10.1007/s10994-009-5119-5>
- Henderi, H. (2021). Comparison of Min-Max normalization and Z-Score Normalization in the K-nearest neighbor (kNN) Algorithm to Test the Accuracy of Types of Breast Cancer. *IJIIS: International Journal of Informatics and Information Systems*, 4(1), 13–20. <https://doi.org/10.47738/ijiis.v4i1.73>
- Herrera, G. P., Constantino, M., Su, J.-J., & Naranpanawa, A. (2023). The use of ICTs and income distribution in Brazil: A machine learning explanation using SHAP values. *Telecommunications Policy*, 47(8), 102598. <https://doi.org/10.1016/j.telpol.2023.102598>
- INEC. (2023). *Metodología Encuesta Estructural Empresarial (ENESEM 2022)*.
- ITU. (2025). *Statistics*. <https://www.itu.int/en/ITU-D/Statistics/pages/stat/default.aspx>
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2023). *An introduction to statistical learning: With applications in R* (Second). Springer Nature.
- Jijo, B. T., & Abdulazeez, A. M. (2021). Classification Based on Decision Tree Algorithm for Machine Learning. *Journal of Applied Science and Technology Trends*, 2(01 SE-Standard Journal Issues), 20–28. <https://doi.org/10.38094/jastt20165>
- Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. *Ijcai*, 14(2), 1137–1145.
- Koutroumpis, P., Leiponen, A., & Thomas, L. D. W. (2020). Small is big in ICT: The impact of R&D on productivity. *Telecommunications Policy*, 44(1), 101833. <https://doi.org/10.1016/j.telpol.2019.101833>
- Kramer, O. (2013). *K-Nearest Neighbors BT - Dimensionality Reduction with Unsupervised Nearest Neighbors* (O. Kramer (ed.); pp. 13–23). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-38652-7_2
- Kuhn, M., & Johnson, K. (2013). Remedies for Severe Class Imbalance. In *Applied*

- Predictive Modeling* (pp. 419–443). Springer New York. https://doi.org/10.1007/978-1-4614-6849-3_16
- Kumari, S., & Singh, S. K. (2022). Deep Learning-based Time Series Models for GDP and ICT Growth Prediction in India. *2022 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS)*, 250–256. <https://doi.org/10.1109/ICCCIS56430.2022.10037636>
- Lee, G., & Xia, W. (2006). Organizational size and IT innovation adoption: A meta-analysis. *Information & Management*, 43(8), 975–985. <https://doi.org/10.1016/j.im.2006.09.003>
- Lee, S., Nam, Y., Lee, S., & Son, H. (2016). Determinants of ICT innovations: A cross-country empirical study. *Technological Forecasting and Social Change*, 110, 71–77. <https://doi.org/10.1016/j.techfore.2015.11.010>
- Lewin, C., & McNicol, S. (2015). Supporting the development of 21st century skills through ICT. *KEYCIT 2014-Key Competencies in Informatics and ICT*, 7, 181–198.
- Luo, Y., & Bu, J. (2016). How valuable is information and communication technology? A study of emerging economy enterprises. *Journal of World Business*, 51(2), 200–211. <https://doi.org/10.1016/j.jwb.2015.06.001>
- Mangeaud, A., Cugno, H., Molina, M. G., Panigo, D. E., Lozano, N. A., Ramello, M. S., & Gonzalez, A. (2024). ¿Cómo asociar variables cualitativas? Un procedimiento ordenado para utilizar pruebas Chi cuadrado de independencia y Regresión Logística. *Revista Methodo: Investigación Aplicada a Las Ciencias Biológicas*, 9(3), 38–47.
- Mehta, M., & Sinha, R. (2022). Women Entrepreneurs and Information Communication Technology: The Journey from Intention to Usage. *Journal of Entrepreneurship and Innovation in Emerging Economies*, 8(2), 228–243. <https://doi.org/10.1177/23939575221089142>
- Meinshausen, N., & Bühlmann, P. (2010). Stability Selection. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 72(4), 417–473. <https://doi.org/10.1111/j.1467-9868.2010.00740.x>
- Menardi, G., & Torelli, N. (2014). Training and assessing classification rules with imbalanced data. *Data Mining and Knowledge Discovery*, 28(1), 92–122. <https://doi.org/10.1007/s10618-012-0295-5>
- Meyer, D., & Wien, F. T. (2001). Support vector machines. *R News*, 1(3), 23–26.
- Molla, A., Gekara, V., Karanasios, S., & Snell, D. (2024). Modeling digital skills beyond the IT workforce: construct definition, measurement and impact on digitalization value.

- Information Technology & People*. <https://doi.org/10.1108/ITP-04-2023-0385>
- Murphy, G., & Siedschlag, I. (2013). Human Capital and Growth of Information and Communication Technology-intensive Industries: Empirical Evidence from Open Economies. *Regional Studies*, 47(9), 1403–1424. <https://doi.org/10.1080/00343404.2010.529115>
- Natekin, A., & Knoll, A. (2013). Gradient boosting machines, a tutorial. *Frontiers in Neurorobotics*, 7, 21. <https://doi.org/10.3389/fnbot.2013.00021>
- Navada, A., Ansari, A. N., Patil, S., & Sonkamble, B. A. (2011). Overview of use of decision tree algorithms in machine learning. *2011 IEEE Control and System Graduate Research Colloquium*, 37–42. <https://doi.org/10.1109/ICSGRC.2011.5991826>
- Neirotti, P., & Pesce, D. (2019). ICT-based innovation and its competitive outcome: the role of information intensity. *European Journal of Innovation Management*, 22(2), 383–404. <https://doi.org/10.1108/EJIM-02-2018-0039>
- Nepelski, D., & De Prato, G. (2014). European ICT Poles of Excellence. *Intereconomics*, 49(6), 324–331. <https://doi.org/10.1007/s10272-014-0515-x>
- Nguyen, Q. H., Ly, H.-B., Ho, L. S., Al-Ansari, N., Le, H. Van, Tran, V. Q., Prakash, I., & Pham, B. T. (2021). Influence of Data Splitting on Performance of Machine Learning Models in Prediction of Shear Strength of Soil. *Mathematical Problems in Engineering*, 2021, 1–15. <https://doi.org/10.1155/2021/4832864>
- Nukala, V. B., & Prasada Rao, S. S. (2021). Role of debt-to-equity ratio in project investment valuation, assessing risk and return in capital markets. *Future Business Journal*, 7(1), 13. <https://doi.org/10.1186/s43093-021-00058-9>
- Orser, B. J., & Riding, A. (2018). The influence of gender on the adoption of technology among SMEs. *International Journal of Entrepreneurship and Small Business*, 33(4), 514. <https://doi.org/10.1504/IJESB.2018.090341>
- Osman, H., Ghafari, M., & Nierstrasz, O. (2017). Automatic feature selection by regularization to improve bug prediction accuracy. *2017 IEEE Workshop on Machine Learning Techniques for Software Quality Evaluation (MaLTeSQuE)*, 27–32. <https://doi.org/10.1109/MALTESQUE.2017.7882013>
- Osunmadewa, K., Folake, O., & Oduroye, A. P. (2024). Benefits of Information Technology to the Development of Nigeria. *International Journal of Mathematics and Computer Research*, 12(10). <https://doi.org/10.47191/ijmcr/v12i10.03>
- Oviedo Gil, Y. M., & Cala Vitery, F. E. (2023). Teleworking: An approach to its

- measurement using official labor surveys in Latin America. *Population Review*, 62(2), 35–59. <https://doi.org/10.1353/prv.2023.a906619>
- Patro, C. S., & Raghunath, K. M. K. (2021). *Digital Transformation* (pp. 16–31). <https://doi.org/10.4018/978-1-7998-2402-2.ch002>
- Piabuo, S. M., Piendiah, N. E., Njamnshi, N. L., & Tieguhong, P. J. (2017). The impact of ICT on the efficiency of HRM in Cameroonian enterprises: Case of the Mobile telephone industry. *Journal of Global Entrepreneurship Research*, 7(1), 7. <https://doi.org/10.1186/s40497-017-0063-5>
- Rabanal, C. (2024). La paradoja de la productividad y el uso de internet en países de América Latina. *Problemas Del Desarrollo. Revista Latinoamericana de Economía*, 55(219), 113–138. <https://doi.org/10.22201/iiec.20078951e.2024.219.70183>
- Rigatti, S. J. (2017). Random Forest. *Journal of Insurance Medicine*, 47(1), 31–39. <https://doi.org/10.17849/in-sm-47-01-31-39.1>
- Rocha, E. M., Pacheco, D. A. de J., Silvério, N., Zanuzzi, C. M. da S., & Selig, P. M. (2024). The role of ICTs for knowledge sharing in franchising networks. *Journal of Business & Industrial Marketing*, 39(2), 366–387. <https://doi.org/10.1108/JBIM-01-2023-0010>
- Selamat, N. A., Abdullah, A., & Mat Diah, N. (2022). Association features of smote and rose for drug addiction relapse risk. *Journal of King Saud University - Computer and Information Sciences*, 34(9), 7710–7719. <https://doi.org/10.1016/j.jksuci.2022.06.012>
- Singh, D., & Singh, B. (2020). Investigating the impact of data normalization on classification performance. *Applied Soft Computing*, 97, 105524. <https://doi.org/10.1016/j.asoc.2019.105524>
- Sørensen, J. B., & Stuart, T. E. (2000). Aging, Obsolescence, and Organizational Innovation. *Administrative Science Quarterly*, 45(1), 81–112. <https://doi.org/10.2307/2666980>
- Suthaharan, S. (2016). *Support Vector Machine BT - Machine Learning Models and Algorithms for Big Data Classification: Thinking with Examples for Effective Learning* (S. Suthaharan (ed.); pp. 207–235). Springer US. https://doi.org/10.1007/978-1-4899-7641-3_9
- Tobar-Pesántez, L., & Solano-Gallegos, S. (2017). La Concentración Financiera y Los Grupos De Poder: Evidencias En La Ciudad De Cuenca-Ecuador (Financial Concentration and Power Groups: Evidence from City of Cuenca-Ecuador). *Revista Internacional Administración & Finanzas*, 10(3), 81–90.
- Touzani, S., Granderson, J., & Fernandes, S. (2018). Gradient boosting machine for

- modeling the energy consumption of commercial buildings. *Energy and Buildings*, 158, 1533–1543. <https://doi.org/https://doi.org/10.1016/j.enbuild.2017.11.039>
- Tripoliti, E. E., Fotiadis, D. I., & Manis, G. (2013). Modifications of the construction and voting mechanisms of the Random Forests Algorithm. *Data & Knowledge Engineering*, 87, 41–65. <https://doi.org/10.1016/j.datak.2013.07.002>
- Vapnik, V., Golowich, S., & Smola, A. (1996). Support vector method for function approximation, regression estimation and signal processing. *Advances in Neural Information Processing Systems*, 9.
- Vas, Z., Szakálné Kanó, I., & Vida, G. (2024). Spatial concentration of the ICT sector in the digital age in Central and Eastern Europe. *European Planning Studies*, 32(12), 2619–2640. <https://doi.org/10.1080/09654313.2024.2396485>
- Verma, C., & Illes, Z. (2019). *Attitude Prediction Towards Ict and Mobile Technology for the Real-time: an Experimental Study Using Machine Learning*. 247–254. <https://doi.org/10.12753/2066-026X-19-171>
- Wadekar, A. S., & Reiter, J. P. (2024). Evaluating Binary Outcome Classifiers Estimated from Survey Data. *Epidemiology*, 35(6), 805–812.
- Wang, H., Ning, H., Lin, Y., Wang, W., Dhelim, S., Farha, F., Ding, J., & Daneshmand, M. (2023). A Survey on the Metaverse: The State-of-the-Art, Technologies, Applications, and Challenges. *IEEE Internet of Things Journal*, 10(16), 14671–14688. <https://doi.org/10.1109/JIOT.2023.3278329>
- Wang, S., & Zhang, H. (2025). Enhancing SMEs sustainable innovation and performance through digital transformation: Insights from strategic technology, organizational dynamics, and environmental adaptation. *Socio-Economic Planning Sciences*, 98, 102124. <https://doi.org/10.1016/j.seps.2024.102124>
- Winship, C., & Radbill, L. (1994). Sampling Weights and Regression Analysis. *Sociological Methods & Research*, 23(2), 230–257. <https://doi.org/10.1177/0049124194023002004>
- Wu, T., & Bajwa, W. U. (2015). Learning the Nonlinear Geometry of High-Dimensional Data: Models and Algorithms. *IEEE Transactions on Signal Processing*, 63(23), 6229–6244. <https://doi.org/10.1109/TSP.2015.2469637>
- Xue, F., Zhao, X., & Tan, Y. (2022). Digital Transformation of Manufacturing Enterprises: An Empirical Study on the Relationships between Digital Transformation, Boundary Spanning, and Sustainable Competitive Advantage. *Discrete Dynamics in Nature and Society*, 2022(1). <https://doi.org/10.1155/2022/4104314>

- Yamada, M., Tang, J., Lugo-Martinez, J., Hodzic, E., Shrestha, R., Saha, A., Ouyang, H., Yin, D., Mamitsuka, H., Sahinalp, C., Radivojac, P., Menczer, F., & Chang, Y. (2018). Ultra High-Dimensional Nonlinear Feature Selection for Big Biological Data. *IEEE Transactions on Knowledge and Data Engineering*, 30(7), 1352–1365. <https://doi.org/10.1109/TKDE.2018.2789451>
- Youden, W. J. (1950). Index for rating diagnostic tests. *Cancer*, 3(1), 32–35. [https://doi.org/10.1002/1097-0142\(1950\)3:1<32::AID-CNCR2820030106>3.0.CO;2-3](https://doi.org/10.1002/1097-0142(1950)3:1<32::AID-CNCR2820030106>3.0.CO;2-3)
- Yunis, M., Tarhini, A., & Kassar, A. (2018). The role of ICT and innovation in enhancing organizational performance: The catalysing effect of corporate entrepreneurship. *Journal of Business Research*, 88, 344–356. <https://doi.org/10.1016/j.jbusres.2017.12.030>
- Zaidi, A. (2022). Mathematical justification on the origin of the sigmoid in logistic regression. *Central European Management Journal*, 30(4), 1327–1337.
- Zhang, H., Crowley, J., Sox, H. C., & A. Olshen, R. (2014). Tree-Structured Statistical Methods. In *Wiley StatsRef: Statistics Reference Online*. Wiley. <https://doi.org/10.1002/9781118445112.stat05678>
- Zhang, L., Zhang, S., & Guo, Y. (2019). The effects of equity financing and debt financing on technological innovation. *Baltic Journal of Management*, 14(4), 698–715. <https://doi.org/10.1108/BJM-01-2019-0011>
- Zhou, Z.-H. (2021). *Machine Learning*. Springer Singapore. <https://doi.org/10.1007/978-981-15-1967-3>
- Živković, L., Štrbac, D., & Paunović, M. (2024). Digitalisation, Growth Vision and Gender Equality Practices in the Machines and Equipment Sector – Does Gender Matter? *JWEE*. <https://doi.org/10.28934/jwee24.12.pp157-176>

ANEXOS

Anexo A. Tablas de predictores

Tabla 13 - Predictores obtenidos mediante la estrategia de regularización (sin categóricas multinivel, particiones = 10 y 5).

<u>No.</u>	<u>Código</u>	<u>Descripción</u>
1	Innovación	Indica si la empresa realizó actividades de investigación y desarrollo orientados a la generación de productos/procesos nuevos o significativamente mejorados.
2	PA_Especialista_TIC	Indica si tiene personal afiliado especialista TIC.
3	Sitio_Web	Indica si la empresa disponía de sitio o página web.
4	Año_compra_software	Indica el año de compra de software y bases de datos.
5	Red_LAN	Indica si la empresa con una red de área local (LAN).
6	Uso_Internet_Capacitacion	Indica si usa el internet para capacitar al personal.
7	MC_CorreoElectronico	Indica si la empresa utilizó como medio de comunicación el correo electrónico.
8	Certificado_Digital	Indica si la página web de la empresa tenía certificado digital.
9	Año_compra_afi	Indica el año de compra de otros activos fijos intangibles.
10	Uso_Internet_Informacion	Indica si usa el internet para obtener información acerca de bienes y servicios (se excluyen los bienes y servicios del sector público).
11	Uso_TIC_RRHH	Indica si la empresa se apoya en el uso de TIC para la gestión de recursos humanos.

12	MC_MediosImpresos	Indica si la empresa utilizó como medio de comunicación los medios impresos.
13	Compras_por_Internet	Indica si la empresa compró bienes o servicios por internet.
14	Compras_por_Internet_porcentaje	Indica el porcentaje de compras por internet.
15	Uso_Internet_Contratacion	Indica si usa el internet para brindar información para contratación de personal.
16	PaginaInterna	Indica si la empresa contaba con una página interna en la Web (intranet/extranet)
17	Uso_Internet_Nube	Indica si usa el internet para servicios en la nube.
18	Energia_Alternativa	Indica si la empresa generó energía eléctrica alternativa o complementaria a la energía de la red pública.
19	Organizaciones_Consultores	Indica si el tipo de organizaciones que cooperaron con la empresa para el desarrollo de actividades de innovación de producto o proceso son los consultores.
20	Uso_TIC_Conocimiento	Indica si la empresa se apoya en el uso de TIC para la gestión del conocimiento.
21	Año_compra_marcas	Indica el año de compra de marcas.
22	Interaccion_EG_GestionElectronica	Indica si la interacción con entidades gubernamentales fue para realizar gestión electrónica completa.
23	Salarios_Nov_ApoyoAd_Mujeres	Indica el monto de sueldos y salarios de personal de apoyo administrativo mujeres en el mes de noviembre.
24	Credito_empleados	Indica si la empresa ofreció créditos a sus empleados.
25	PA_ApoyoAd_Mujeres	Indica el total de personal de apoyo administrativo mujer.

26	Interaccion_EG_ConseguirFormularios	Indica si la interacción con entidades gubernamentales fue para conseguir formularios.
27	Año_RUC	Indica desde qué año la empresa ha tenido un RUC.
28	Valor_elaboracion_CP_marcas	Indica el valor de elaboración por cuenta propia de Marcas.
29	Salarios_Nov_Mujeres	Indica el monto de sueldos y salarios de mujeres en el mes de noviembre.
30	Salarios_Mujeres	Indica el monto de sueldos y salarios de mujeres en el año.
31	SoftwareAbierto_AplicacionesOfimaticas	Indica si la empresa utilizó software abierto: aplicaciones ofimáticas.
32	Edad	Indica la edad de la empresa

Nota: En ambos esquemas de validación (5 y 10), el conjunto de predictores seleccionados fue idéntico, por lo que se muestra una única tabla.

Tabla 14 - Predictores obtenidos mediante la estrategia de regularización (con categóricas multinivel, partición = 10).

<u>No.</u>	<u>Código</u>	<u>Descripción</u>
1	Innovación	Indica si la empresa realizó actividades de investigación y desarrollo orientados a la generación de productos/procesos nuevos o significativamente mejorados.
2	PA_Especialista_TIC	Indica si tiene personal afiliado especialista TIC.
3	Año_compra_software	Indica el año de compra de software y bases de datos.
4	Uso_Internet_Capacitacion	Indica si usa el internet para capacitar al personal.
5	Sitio_Web	Indica si la empresa disponía de sitio o página web.

6	Red_LAN	Indica si la empresa con una red de área local (LAN).
7	Certificado_Digital	Indica si la página web de la empresa tenía certificado digital.
8	Compras_por_Internet_porcentaje	Indica el porcentaje de compras por internet.
9	Año_compra_afi	Indica el año de compra de otros activos fijos intangibles.
10	Empresa_Grande	Indica si la empresa es Grande.
11	MC_MediosImpresos	Indica si la empresa utilizó como medio de comunicación los medios impresos.
12	Uso_Internet_Contratacion	Indica si usa el internet para brindar información para contratación de personal.
13	Cambio_tamano	Indica si la empresa cambió de tamaño en comparación al año anterior.
14	PaginaInterna	Indica si la empresa contaba con una página interna en la Web (intranet/extranet)
15	Energia_Alternativa	Indica si la empresa generó energía eléctrica alternativa o complementaria a la energía de la red pública.
16	Uso_TIC_RRHH	Indica si la empresa se apoya en el uso de TIC para la gestión de recursos humanos.
17	Empresa_MedianaB	Indica si la empresa es Mediana B.
18	Region_Costa	Indica si la empresa pertenece a la región Costa.
19	Año_compra_marcas	Indica el año de compra de marcas.
20	Valor_elaboracion_CP_marcas	Indica el valor de elaboración por cuenta propia de Marcas.
21	Salarios_Nov_ApoyoAd_Mujeres	Indica el monto de sueldos y salarios de personal de apoyo administrativo mujeres en el mes de noviembre.
22	Credito_empleados	Indica si la empresa ofreció créditos a sus empleados.

23	Uso_Internet_Informacion	Indica si usa el internet para obtener información acerca de bienes y servicios (se excluyen los bienes y servicios del sector público).
24	Consumo_Gasolina_galones	Indica la cantidad consumida de Gasolina Extra (Galones).
25	Costo_viaje	Indica el costo de viaje.
26	Año_elaboracion_CP_afi	Indica el año de elaboración por cuenta propia de otros activos fijos intangibles.
27	Uso_TIC_Conocimiento	Indica si la empresa se apoya en el uso de TIC para la gestión del conocimiento.
28	Venta_AF_edificios	Indica el monto de venta de edificios.
29	Utilidad_venta_AF_Maquinaria	Indica la utilidad por venta de maquinaria, equipo e Instalaciones.
30	Ganancias_netas_actividades_dis continuadas	Indica ganancias netas procedentes de actividades discontinuadas.
31	Credito_banco_privado	Indica si el tipo de entidad financiera que otorgó el crédito es un Banco Privado.
32	CS_nacional_privado	Indica el porcentaje de capital social nacional privado.
33	Cantidad_Agua_Consumida	Indica la cantidad de agua consumida.
34	Compras_por_Internet_nacional	Indica el porcentaje de ventas por internet que se realizó en el país.
35	Año_elaboración_CP_derechoLl ave	Indica el año de elaboración por cuenta propia de derechos de llave.
36	Endeudamiento_Patrimonio	Indica el ratio del nivel de endeudamiento del patrimonio.
37	Año_RUC	Indica desde qué año la empresa ha tenido un RUC.
38	Consumo_Combustible	Indica si la empresa consumió combustibles y/o lubricantes en diferentes procesos.
39	Adq_AFU_EquipoCamionero	Indica el monto de adquisición de equipo camionero móvil usados.

40	Año_compra_derechoLlave	Indica el año de compra de derechos de llave.
41	Organizaciones_Consultores	Indica si el tipo de organizaciones que cooperaron con la empresa para el desarrollo de actividades de innovación de producto o proceso son los consultores.
42	CS_extranjero_publico	Indica el porcentaje de capital social extranjero público.
43	Regalias_cesiones	Indica el ingreso por regalías y otras cesiones de derechos.
44	Costos_impuestos	Indica el costo de impuestos, contribuciones y otros.
45	Gastos_seguros	Indica el gasto de seguros y reaseguros (Primas y Cesiones).
46	Utilidad_venta_AF_Terrenos	Indica la utilidad por venta de terrenos.
47	Salarios_Nov_Directores_Hombres	Indica el monto de sueldos y salarios de directores y gerentes hombres en el mes de noviembre.
48	Adq_AFU_Maquinaria	Indica el monto de adquisición de maquinaria en montaje y en tránsito usado.
49	Const_AF_CP_Edificios	Indica el monto por construcciones de edificios.
50	Venta_AF_Maquinarias	Indica el monto de venta de maquinaria, equipo e Instalaciones.
51	Adq_Muebles	Indica el monto en adquisición de muebles y enseres y equipos de oficina nuevos.
52	No_Cambio_tamano	Indica si la empresa no cambió de tamaño en comparación al año anterior.
53	Gastos_indirectos_exterior	Indica gastos indirectos asignados desde el exterior por partes relacionadas.
54	Adq_AFN_Vehiculos	Indica el monto de adquisición de vehículos y equipo de transporte nuevos.
55	Adq_AFU_EquipoComputacion	Indica el monto de adquisición de equipos de computación usados.

56	Obst_VI_Caracteristicas	Indica si entre los obstáculos para vender por internet es la características del bien o servicio.
57	Credito_otros	Indica si otro tipo de entidad financiera otorgó el crédito.
58	Compras_por_Internet_extranjero	Indica el porcentaje de ventas por internet que se realizó en el extranjero.
59	Dispositivos_PDA	Indica el número de dispositivos PDA.
60	Costos_ajustes	Indica los costos de ajustes.
61	Valor_produccion_Energia_Solar	Indica el valor de producción de energía solar.
62	Salarios_Nov_Operadores_Mujeres	Indica el monto de sueldos y salarios de operadores mujeres en el mes de noviembre.
63	Adq_Vehiculos	Indica el monto en adquisición de vehículos y equipos de transporte nuevos.
64	Revalorizaciones_Terrenos	Indica las revalorizaciones y ajustes por el valor de mercado por concepto de terrenos.
65	Dep_EquipoCamionero	Indica la depreciación acumulada de equipo camionero móvil.
66	Costo_PromocionPublicidad	Indica el costo en promoción y publicidad.
67	Uso_Internet_Nube	Indica si usa el internet para servicios en la nube.
68	Edad	Indica la edad de la empresa
69	Salarios_Nov_TNC	Indica el monto de sueldos y salarios de trabajadores no calificados total en el mes de noviembre.
70	Venta_energia_BiomasaProducida	Indica la cantidad (KWh/año) de venta de energía por biomasa producida.
71	Año_elaboracion_CP_software	Indica el año de elaboración por cuenta propia de software y bases de datos.
72	Intereses_IF	Indica el gasto por intereses con instituciones financieras.
73	Duracion_AC_franquicia	Indica la duración (en años) de franquicias compradas.
74	Costo_transporte	Indica el costo en transporte.

75	CS_nacional_publico	Indica el porcentaje de capital social nacional público.
76	Duracion_AC_AFI	Indica la duración (en años) de otros activos fijos intangibles comprados.
77	Costos_gastos_Operaciones	Indica el total de Costos y Gastos de Operaciones de regalías, servicios técnicos, administrativos, de consultoría y similares.
78	ID_Externa	Indica el valor que destinó la empresa a Investigación y Desarrollo (I+D) externa.
79	Año_compra_patentes	Indica el año de compra de patentes.
80	Region_Amazonia	Indica si la empresa pertenece a la región Amazonía.
81	Utilidad_venta_AF_Abiologicos	Indica la utilidad por venta de activos biológicos
82	PA_ApoyoAd_Mujeres	Indica el total de personal de apoyo administrativo mujer.
83	Duracion_AE_CP_marcas	Indica la duración (en años) de marcas elaboradas por cuenta propia.
84	Valor_GasolinaSuper	Indica el monto de gasolina súper consumida.
85	Costo_Honorarios	Indica el costo por honorarios profesionales y dietas.
86	Region_Insular	Indica si la empresa pertenece a la región Insular.
87	Construcciones_AF_CP_OtrosA FI	Indica el monto por construcciones de otros activos fijos tangibles.
88	Duracion_AE_CP_derechosLlav e	Indica la duración (en años) de derechos de llave elaborados por cuenta propia.

Tabla 15 - Predictores obtenidos mediante la estrategia de regularización (con categóricas multinivel, partición = 5).

<u>No.</u>	<u>Código</u>	<u>Descripción</u>
------------	---------------	--------------------

1	Innovación	Indica si la empresa realizó actividades de investigación y desarrollo orientados a la generación de productos/procesos nuevos o significativamente mejorados.
2	PA_Especialista_TIC	Indica si tiene personal afiliado especialista TIC.
3	Año_compra_software	Indica el año de compra de software y bases de datos.
4	Uso_Internet_Capacitacion	Indica si usa el internet para capacitar al personal.
5	Sitio_Web	Indica si la empresa disponía de sitio o página web.
6	Red_LAN	Indica si la empresa con una red de área local (LAN).
7	Certificado_Digital	Indica si la página web de la empresa tenía certificado digital.
8	Compras_por_Internet_porcentaje	Indica el porcentaje de compras por internet.
9	Año_compra_afi	Indica el año de compra de otros activos fijos intangibles.
10	Empresa_Grande	Indica si la empresa es Grande.
11	Uso_Internet_Contratacion	Indica si usa el internet para brindar información para contratación de personal.
12	MC_MediosImpresos	Indica si la empresa utilizó como medio de comunicación los medios impresos.
13	Energia_Alternativa	Indica si la empresa generó energía eléctrica alternativa o complementaria a la energía de la red pública.
14	PaginaInterna	Indica si la empresa contaba con una página interna en la Web (intranet/extranet)
15	Empresa_MedianaB	Indica si la empresa es Mediana B.
16	Uso_TIC_RRHH	Indica si la empresa se apoya en el uso de TIC para la gestión de recursos humanos.

17	Cambio_tamano	Indica si la empresa cambió de tamaño en comparación al año anterior.
18	Año_compra_marcas	Indica el año de compra de marcas.
19	Valor_elaboracion_CP_marcas	Indica el valor de elaboración por cuenta propia de Marcas.
20	Credito_empleados	Indica si la empresa ofreció créditos a sus empleados.
21	Salarios_Nov_ApoyoAd_Mujeres	Indica el monto de sueldos y salarios de personal de apoyo administrativo mujeres en el mes de noviembre.
22	Consumo_Gasolina_galones	Indica la cantidad consumida de Gasolina Extra (Galones).
23	Uso_TIC_Conocimiento	Indica si la empresa se apoya en el uso de TIC para la gestión del conocimiento.
24	Año_elaboracion_CP_afis	Indica el año de elaboración por cuenta propia de otros activos fijos intangibles.
25	Año_RUC	Indica desde qué año la empresa ha tenido un RUC.
26	Compras_por_Internet_nacional	Indica el porcentaje de ventas por internet que se realizó en el país.
27	Costo_viaje	Indica el costo de viaje.
28	Venta_AF_edificios	Indica el monto de venta de edificios.
29	CS_nacional_privado	Indica el porcentaje de capital social nacional privado.
30	Credito_banco_privado	Indica si el tipo de entidad financiera que otorgó el crédito es un Banco Privado.
31	Cantidad_Agua_Consumida	Indica la cantidad de agua consumida.
32	Ganancias_netas_actividades_discontinuas	Indica ganancias netas procedentes de actividades discontinuadas.
33	Utilidad_venta_AF_Maquinaria	Indica la utilidad por venta de maquinaria, equipo e Instalaciones.
34	Año_elaboración_CP_derechos de llave	Indica el año de elaboración por cuenta propia de derechos de llave.

35	Uso_Internet_Informacion	Indica si usa el internet para obtener información acerca de bienes y servicios (se excluyen los bienes y servicios del sector público).
36	Consumo_Combustible	Indica si la empresa consumió combustibles y/o lubricantes en diferentes procesos.
37	Region_Costa	Indica si la empresa pertenece a la región Costa.
38	Endeudamiento_Patrimonio	Indica el ratio del nivel de endeudamiento del patrimonio.
39	Año_compra_derechoLlave	Indica el año de compra de derechos de llave.
40	Organizaciones_Consultores	Indica si el tipo de organizaciones que cooperaron con la empresa para el desarrollo de actividades de innovación de producto o proceso son los consultores.
41	Adq_AFN_Vehiculos	Indica el monto de adquisición de vehículos y equipo de transporte nuevos.
42	Gastos_seguros	Indica el gasto de seguros y reaseguros (Primas y Cesiones).
43	Adq_AFU_EquipoCamionero	Indica el monto de adquisición de equipo caminonero móvil usados.
44	Const_AF_CP_Edificios	Indica el monto por construcciones de edificios.
45	CS_extranjero_publico	Indica el porcentaje de capital social extranjero público.
46	Utilidad_venta_AF_Terrenos	Indica la utilidad por venta de terrenos.
47	Adq_AFU_Maquinaria	Indica el monto de adquisición de maquinaria en montaje y en tránsito usado.
48	Compras_por_Internet_extranjero o	Indica el porcentaje de ventas por internet que se realizó en el extranjero.
49	Regalias_cesiones	Indica el ingreso por regalías y otras cesiones de derechos.
50	No_Cambio_tamano	Indica si la empresa no cambió de tamaño en comparación al año anterior.

51	Venta_AF_Maquinarias	Indica el monto de venta de maquinaria, equipo e Instalaciones.
52	Adq_AFN_Muebles	Indica el monto de adquisición de muebles y enseres y equipos de Oficina nuevos.
53	Adq_AFU_EquipoComputacion	Indica el monto de adquisición de equipos de computación usados.
54	Credito_otros	Indica si otro tipo de entidad financiera otorgó el crédito.
55	Salarios_Nov_Operadores_Mujeres	Indica el monto de sueldos y salarios de operadores mujeres en el mes de noviembre.
56	Dispositivos_PDA	Indica el número de dispositivos PDA.
57	Costos_impuestos	Indica el costo de impuestos, contribuciones y otros.
58	Gastos_indirectos_exterior	Indica gastos indirectos asignados desde el exterior por partes relacionadas.
59	PA_ApoyoAd_Mujeres	Indica el total de personal de apoyo administrativo mujer.
60	Costos_ajustes	Indica los costos de ajustes.
61	Edad	Indica la edad de la empresa
62	Valor_produccion_Energia_Solar	Indica el valor de producción de energía solar.
63	Salarios_Nov_Directores_Hombres	Indica el monto de sueldos y salarios de directores y gerentes hombres en el mes de noviembre.
64	Salarios_Nov_TNC	Indica el monto de sueldos y salarios de trabajadores no calificados total en el mes de noviembre.
65	Dep_EquipoCamionero	Indica la depreciación acumulada de equipo camionero móvil.
66	Año_elaboracion_CP_software	Indica el año de elaboración por cuenta propia de software y bases de datos.
67	Revalorizaciones_Terrenos	Indica las revalorizaciones y ajustes por el valor de mercado por concepto de terrenos.

68	Adq_Muebles	Indica el monto en adquisición de muebles y enseres y equipos de oficina nuevos.
69	Region_Amazonia	Indica si la empresa pertenece a la región Amazonía.
70	Año_compra_patentes	Indica el año de compra de patentes.
71	Adq_Vehiculos	Indica el monto en adquisición de vehículos y equipos de transporte nuevos.
72	Duracion_AC_AFI	Indica la duración (en años) de otros activos fijos intangibles comprados.
73	Region_Insular	Indica si la empresa pertenece a la región Insular.
74	Venta_energia_BiomasaProducida	Indica la cantidad (KWh/año) de venta de energía por biomasa producida.
75	CS_nacional_publico	Indica el porcentaje de capital social nacional público.
76	Costo_PromocionPublicidad	Indica el costo en promoción y publicidad.
77	Duracion_AE_CP_derechosLlave	Indica la duración (en años) de derechos de llave elaborados por cuenta propia.

Tabla 16 - Predictores obtenidos mediante la estrategia híbrida (con categóricas multinivel, partición = 10).

<u>No.</u>	<u>Código</u>	<u>Descripción</u>
1	Innovación	Indica si la empresa realizó actividades de investigación y desarrollo orientados a la generación de productos/procesos nuevos o significativamente mejorados.
2	PA_Especialista_TIC	Indica si tiene personal afiliado especialista TIC.
3	Sitio_Web	Indica si la empresa disponía de sitio o página web.
4	Compras_por_Internet	Indica si la empresa compró bienes o servicios por internet.

5	Region_Costa	Indica si la empresa pertenece a la región Costa.
6	Total_PA_Mujeres	Indica el total de mujeres.
7	Edad	Indica la edad de la empresa
8	ID_Externa	Indica el valor que destinó la empresa a Investigación y Desarrollo (I+D) externa.
9	Total_Dispositivos	Indica el total de dispositivos.
10	Endeudamiento_Patrimonio	Indica el ratio del nivel de endeudamiento del patrimonio.
11	VAE	Indica el valor agregado empresarial.
12	Region_Amazonia	Indica si la empresa pertenece a la región Amazonía.
13	PA_UsaInternet_Mujeres	Indica el personal afiliado hombre que utiliza internet.
14	ROE	Indica el ratio del retorno sobre el patrimonio (ROE)
15	Total_PA	Indica el total del personal afiliado.
16	Region_Insular	Indica si la empresa pertenece a la región Insular.
17	FirmaDigital	Indica si la empresa usó firma digital en comunicaciones enviadas.

Tabla 17 - Predictores obtenidos mediante la estrategia híbrida (con categóricas multinivel, partición = 5).

<u>No.</u>	<u>Código</u>	<u>Descripción</u>
1	Innovación	Indica si la empresa realizó actividades de investigación y desarrollo orientados a la generación de productos/procesos nuevos o significativamente mejorados.
2	PA_Especialista_TIC	Indica si tiene personal afiliado especialista TIC.
3	Sitio_Web	Indica si la empresa disponía de sitio o página web.

4	Compras_por_Internet	Indica si la empresa compró bienes o servicios por internet.
5	Region_Costa	Indica si la empresa pertenece a la región Costa.
6	Total_PA_Mujeres	Indica el total de mujeres.
7	Edad	Indica la edad de la empresa
8	ID_Externa	Indica el valor que destinó la empresa a Investigación y Desarrollo (I+D) externa.
9	Total_Dispositivos	Indica el total de dispositivos.
10	Endeudamiento_Patrimonio	Indica el ratio del nivel de endeudamiento del patrimonio.
11	VAE	Indica el valor agregado empresarial.
12	Region_Amazonia	Indica si la empresa pertenece a la región Amazonía.
13	ROE	Indica el ratio del retorno sobre el patrimonio (ROE)
14	PA_UsaInternet_Mujeres	Indica el personal afiliado hombre que utiliza internet.
15	Total_PA	Indica el total del personal afiliado.
16	Region_Insular	Indica si la empresa pertenece a la región Insular.

Tabla 18 - Predictores obtenidos mediante la estrategia híbrida (sin categóricas multinivel, partición = 10).

<u>No.</u>	<u>Código</u>	<u>Descripción</u>
1	Innovación	Indica si la empresa realizó actividades de investigación y desarrollo orientados a la generación de productos/procesos nuevos o significativamente mejorados.
2	PA_Especialista_TIC	Indica si tiene personal afiliado especialista TIC.
3	Sitio_Web	Indica si la empresa disponía de sitio o página web.

4	Internet	Indica si la empresa disponía de conexión a internet.
5	Compras_por_Internet	Indica si la empresa compró bienes o servicios por internet.
6	FirmaDigital	Indica si la empresa usó firma digital en comunicaciones enviadas.
7	Total_PA_Mujeres	Indica el total de mujeres.
8	Total_Dispositivos	Indica el total de dispositivos.
9	Edad	Indica la edad de la empresa
10	VAE	Indica el valor agregado empresarial.

Tabla 19 - Predictores obtenidos mediante la estrategia híbrida (sin categóricas multinivel, partición = 5).

<u>No.</u>	<u>Código</u>	<u>Descripción</u>
1	Innovación	Indica si la empresa realizó actividades de investigación y desarrollo orientados a la generación de productos/procesos nuevos o significativamente mejorados.
2	PA_Especialista_TIC	Indica si tiene personal afiliado especialista TIC.
3	Sitio_Web	Indica si la empresa disponía de sitio o página web.
4	Internet	Indica si la empresa disponía de conexión a internet.
5	Compras_por_Internet	Indica si la empresa compró bienes o servicios por internet.
6	FirmaDigital	Indica si la empresa usó firma digital en comunicaciones enviadas.
7	Total_PA_Mujeres	Indica el total de mujeres.
8	Total_Dispositivos	Indica el total de dispositivos.
9	Edad	Indica la edad de la empresa
10	VAE	Indica el valor agregado empresarial.

11	Endeudamiento_Patrimonio	Indica el ratio del nivel de endeudamiento del patrimonio.
----	--------------------------	--

Anexo B. Tablas de desempeño predictivo

Tabla 20 - Evaluación de modelos a partir de la estrategia de regularización (sin categóricas multinivel, partición = 10).

Modelo	BACC	Precisión	Recuperación	Especificidad	F1 Score	AUC
LR	0.253	0.412	0.311	0.194	0.354	0.813
DT	0.633	0.911	0.324	0.942	0.478	0.364
KNN	0.694	0.852	0.565	0.823	0.679	0.709
RF	0.751	0.856	0.724	0.779	0.784	0.814
GBM	0.744	0.849	0.721	0.767	0.780	0.804
SVM (linear)	0.734	0.865	0.653	0.815	0.745	0.734
SVM (radial)	0.734	0.876	0.628	0.839	0.732	0.734

Tabla 21 - Evaluación de modelos a partir de la estrategia de regularización (sin categóricas multinivel, partición = 5).

Modelo	BACC	Precisión	Recuperación	Especificidad	F1 Score	AUC
LR	0.255	0.431	0.353	0.156	0.388	0.809
DT	0.622	0.917	0.292	0.952	0.443	0.380
KNN	0.693	0.867	0.534	0.851	0.661	0.729
RF	0.745	0.836	0.761	0.729	0.796	0.804
GBM	0.736	0.835	0.735	0.736	0.782	0.788
SVM (linear)	0.746	0.869	0.676	0.815	0.760	0.746
SVM (radial)	0.711	0.911	0.513	0.909	0.657	0.711

Tabla 22 - Evaluación de modelos a partir de la estrategia de regularización (con categóricas multinivel, partición = 10).

Modelo	BACC	Precisión	Recuperación	Especificidad	F1 Score	AUC
LR	0.239	0.404	0.311	0.168	0.351	0.822
DT	0.500	NaN*	0.000	1.000	NaN	0.500
KNN	0.583	0.883	0.220	0.947	0.352	0.416
RF	0.744	0.839	0.749	0.739	0.791	0.807
GBM	0.541	0.970	0.086	0.995	0.158	0.458
SVM (linear)	0.745	0.874	0.663	0.827	0.754	0.745
SVM (radial)	0.575	0.904	0.187	0.964	0.309	0.575

* El valor NaN corresponde a una situación en que el criterio de Youden no pudo aplicarse (única variable de partición). Se adoptó un umbral fijo de 0.5 para permitir la estimación de métricas.

Tabla 23 - Evaluación de modelos a partir de la estrategia de regularización (con categóricas multinivel, partición = 5).

Modelo	BACC	Precisión	Recuperación	Especificidad	F1 Score	AUC
LR	0.241	0.397	0.295	0.187	0.338	0.819
DT	0.500	NaN*	0.000	1.000	NaN	0.500
KNN	0.602	0.857	0.294	0.911	0.437	0.397
RF	0.525	1.000	0.050	1.000	0.096	0.474
GBM	0.525	1.000	0.050	1.000	0.096	0.473
SVM (linear)	0.759	0.880	0.689	0.830	0.773	0.759
SVM (radial)	0.577	0.884	0.202	0.952	0.329	0.577

* El valor NaN corresponde a una situación en que el criterio de Youden no pudo aplicarse (única variable de partición). Se adoptó un umbral fijo de 0.5 para permitir la estimación de métricas.

Tabla 24 - Evaluación de modelos a partir de la estrategia de selección manual (con categóricas multinivel, partición = 10).

Modelo	BACC	Precisión	Recuperación	Especificidad	F1 Score	AUC
LR	0.738	0.843	0.718	0.758	0.776	0.788
DT	0.514	0.857	0.040	0.988	0.076	0.514
KNN	0.629	0.838	0.397	0.861	0.539	0.660
RF	0.640	0.904	0.348	0.933	0.502	0.641
GBM	0.522	0.847	0.066	0.978	0.123	0.533
SVM (linear)	0.714	0.861	0.606	0.823	0.711	0.714
SVM (radial)	0.615	0.886	0.299	0.930	0.447	0.385

Tabla 25 - Evaluación de modelos a partir de la estrategia de selección manual (con categóricas multinivel, partición = 5).

Modelo	BACC	Precisión	Recuperación	Especificidad	F1 Score	AUC
LR	0.738	0.843	0.718	0.758	0.776	0.788
DT	0.514	0.857	0.040	0.988	0.076	0.514
KNN	0.629	0.838	0.397	0.861	0.539	0.660
RF	0.650	0.892	0.384	0.916	0.537	0.660
GBM	0.515	0.929	0.034	0.995	0.066	0.526
SVM (linear)	0.714	0.861	0.606	0.823	0.711	0.714
SVM (radial)	0.615	0.886	0.299	0.930	0.447	0.385

Tabla 26 - Evaluación de modelos a partir de la estrategia de selección manual (sin categóricas multinivel, partición = 10).

Modelo	BACC	Precisión	Recuperación	Especificidad	F1 Score	AUC
LR	0.736	0.866	0.656	0.815	0.746	0.792
DT	0.514	0.857	0.040	0.988	0.076	0.514
KNN	0.642	0.837	0.440	0.844	0.577	0.656
RF	0.691	0.767	0.847	0.535	0.805	0.752
GBM	0.576	0.930	0.176	0.976	0.296	0.586
SVM (linear)	0.715	0.856	0.619	0.811	0.718	0.715
SVM (radial)	0.603	0.899	0.259	0.947	0.402	0.397

Tabla 27 - Evaluación de modelos a partir de la estrategia de selección manual (sin categóricas multinivel, partición = 5).

Modelo	BACC	Precisión	Recuperación	Especificidad	F1 Score	AUC
LR	0.736	0.866	0.656	0.815	0.746	0.792
DT	0.514	0.857	0.040	0.988	0.076	0.514
KNN	0.642	0.837	0.440	0.844	0.577	0.656
RF	0.652	0.891	0.390	0.914	0.543	0.660
GBM	0.515	0.929	0.034	0.995	0.066	0.525
SVM (linear)	0.715	0.856	0.619	0.811	0.718	0.715
SVM (radial)	0.603	0.899	0.259	0.947	0.402	0.397

Tabla 28 - Evaluación de modelos a partir de la estrategia híbrida (con categóricas multinivel, partición = 10).

Modelo	BACC	Precisión	Recuperación	Especificidad	F1 Score	AUC
LR	0.259	0.418	0.316	0.201	0.360	0.790
DT	0.502	0.645	0.999	0.005	0.784	0.498
KNN	0.676	0.865	0.491	0.861	0.626	0.718
RF	0.709	0.849	0.616	0.801	0.714	0.771
GBM	0.515	0.865	0.042	0.988	0.081	0.482
SVM (linear)	0.723	0.870	0.611	0.835	0.718	0.723
SVM (radial)	0.714	0.806	0.758	0.669	0.781	0.714

Tabla 29 - Evaluación de modelos a partir de la estrategia híbrida (con categóricas multinivel, partición = 5).

Modelo	BACC	Precisión	Recuperación	Especificidad	F1 Score	AUC
LR	0.262	0.410	0.296	0.228	0.344	0.788
DT	0.502	0.645	0.999	0.005	0.784	0.498
KNN	0.678	0.808	0.624	0.731	0.704	0.707
RF	0.729	0.860	0.651	0.808	0.741	0.783
GBM	0.558	0.874	0.156	0.959	0.265	0.440
SVM (linear)	0.706	0.852	0.601	0.811	0.704	0.706
SVM (radial)	0.615	0.889	0.298	0.933	0.446	0.615

Tabla 30 - Evaluación de modelos a partir de la estrategia híbrida (sin categóricas multinivel, partición = 10).

Modelo	BACC	Precisión	Recuperación	Especificidad	F1 Score	AUC
LR	0.260	0.416	0.311	0.209	0.356	0.792
DT	0.510	0.649	1.000	0.019	0.787	0.490
KNN	0.607	0.899	0.270	0.945	0.415	0.634
RF	0.703	0.822	0.667	0.739	0.736	0.769
GBM	0.510	0.649	1.000	0.019	0.787	0.490
SVM (linear)	0.718	0.861	0.615	0.820	0.718	0.718
SVM (radial)	0.711	0.856	0.607	0.815	0.711	0.711

Tabla 31 - Evaluación de modelos a partir de la estrategia híbrida (sin categóricas multinivel, partición = 5).

Modelo	BACC	Precisión	Recuperación	Especificidad	F1 Score	AUC
LR	0.258	0.413	0.307	0.209	0.352	0.791
DT	0.510	0.649	1.000	0.019	0.787	0.490
KNN	0.707	0.851	0.606	0.808	0.708	0.757
RF	0.730	0.825	0.747	0.712	0.784	0.783
GBM	0.513	0.651	0.997	0.029	0.787	0.487
SVM (linear)	0.719	0.859	0.622	0.815	0.721	0.719
SVM (radial)	0.697	0.881	0.521	0.873	0.655	0.697

Anexo C. Análisis de Estabilidad

Tabla 32 - Predictores estables de la estrategia de regularización (sin categóricas multinivel, particiones = 10 y 5).

<u>No.</u>	<u>Código</u>	$\hat{\pi}_j$	<u>Descripción</u>
1	Red_LAN	0.98	Indica si la empresa con una red de área local (LAN).
2	PA_Especialista_TIC	0.98	Indica si tiene personal afiliado especialista TIC.
3	Innovación	0.98	Indica si la empresa realizó actividades de investigación y desarrollo orientados a la generación de productos/procesos nuevos o significativamente mejorados.
4	Año_compra_software	0.98	Indica el año de compra de software y bases de datos.
5	Año_compra_afi	0.98	Indica el año de compra de otros activos fijos intangibles.
6	Uso_Internet_Capacitacion	0.97	Indica si usa el internet para capacitar al personal.
7	Compras_por_Internet_porcentaje	0.94	Indica el porcentaje de compras por internet.
8	Energia_Alternativa	0.88	Indica si la empresa generó energía eléctrica alternativa o complementaria a la energía de la red pública.
9	Certificado_Digital	0.85	Indica si la página web de la empresa tenía certificado digital.
10	Sitio_Web	0.81	Indica si la empresa disponía de sitio o página web.
11	Año_compra_marcas	0.79	Indica el año de compra de marcas.
12	PaginaInterna	0.78	Indica si la empresa contaba con una página interna en la Web (intranet/extranet)
13	MC_CorreoElectronico	0.75	Indica si la empresa utilizó como medio de comunicación el correo electrónico.

14	Compras_por_Internet	0.72	Indica si la empresa compró bienes o servicios por internet.
15	MC_MediosImpresos	0.72	Indica si la empresa utilizó como medio de comunicación los medios impresos.
16	Organizaciones_Consultores	0.7	Indica si el tipo de organizaciones que cooperaron con la empresa para el desarrollo de actividades de innovación de producto o proceso son los consultores.

Tabla 33 - Predictores estables de la estrategia de regularización (con categóricas multinivel, partición = 10).

<u>No.</u>	<u>Código</u>	$\hat{\pi}_j$	<u>Descripción</u>
1	Uso_Internet_Capacitacion	1	Indica si usa el internet para capacitar al personal.
2	Red_LAN	1	Indica si la empresa con una red de área local (LAN).
3	Compras_por_Internet_porcentaje	1	Indica el porcentaje de compras por internet.
4	PA_Especialista_TIC	1	Indica si tiene personal afiliado especialista TIC.
5	Innovación	1	Indica si la empresa realizó actividades de investigación y desarrollo orientados a la generación de productos/procesos nuevos o significativamente mejorados.
6	Año_compra_software	1	Indica el año de compra de softwaree y bases de datos.
7	Año_compra_afi	1	Indica el año de compra de otros activos fijos intangibles.
8	Energia_Alternativa	0.97	Indica si la empresa generó energía eléctrica alternativa o complementaria a la energía de la red pública.

9	Empresa_MedianaB	0.89	Indica si la empresa es Mediana B.
10	Empresa_Grande	0.89	Indica si la empresa es Grande.
11	Año_compra_marcas	0.89	Indica el año de compra de marcas.
12	Certificado_Digital	0.88	Indica si la página web de la empresa tenía certificado digital.
13	Sitio_Web	0.85	Indica si la empresa disponía de sitio o página web.
14	Año_RUC	0.84	Indica desde qué año la empresa ha tenido un RUC.
15	Cambio_tamano	0.83	Indica si la empresa cambió de tamaño en comparación al año anterior.
16	No_Cambio_tamano	0.83	Indica si la empresa no cambió de tamaño en comparación al año anterior.
17	Edad	0.83	Indica la edad de la empresa
18	Credito_empleados	0.82	Indica si la empresa ofreció créditos a sus empleados.
19	PaginaInterna	0.81	Indica si la empresa contaba con una página interna en la Web (intranet/extranet)
20	Compras_por_Internet_nacional	0.79	Indica el porcentaje de ventas por internet que se realizó en el país.
21	MC_MediosImpresos	0.78	Indica si la empresa utilizó como medio de comunicación los medios impresos.
22	Region_Costa	0.75	Indica si la empresa pertenece a la región Costa.
23	Region_Amazonia	0.75	Indica si la empresa pertenece a la región Amazonía.
24	Region_Insular	0.75	Indica si la empresa pertenece a la región Insular.
25	Adq_Muebles	0.73	Indica el monto en adquisición de muebles y enseres y equipos de oficina nuevos.
26	Adq_AFN_Muebles	0.73	Indica el monto de adquisición de muebles y enseres y equipos de Oficina nuevos.

27	Valor_elaboracion_CP_marcas	0.73	Indica el valor de elaboración por cuenta propia de Marcas.
28	Costo_viaje	0.72	Indica el costo de viaje.
29	Año_elaboracion_CP_afi	0.72	Indica el año de elaboración por cuenta propia de otros activos fijos intangibles.
30	Adq_Vehiculos	0.71	Indica el monto en adquisición de vehículos y equipos de transporte nuevos.
31	Consumo_Combustible	0.71	Indica si la empresa consumió combustibles y/o lubricantes en diferentes procesos.

Tabla 34 – Predictores estables de la estrategia de regularización (con categóricas multinivel, partición = 5).

<u>No.</u>	<u>Código</u>	$\hat{\pi}_j$	<u>Descripción</u>
1	Uso_Internet_Capacitacion	1	Indica si usa el internet para capacitar al personal.
2	Red_LAN	1	Indica si la empresa con una red de área local (LAN).
3	Compras_por_Internet_porcentaje	1	Indica el porcentaje de compras por internet.
4	PA_Especialista_TIC	1	Indica si tiene personal afiliado especialista TIC.
5	Innovación	1	Indica si la empresa realizó actividades de investigación y desarrollo orientados a la generación de productos/procesos nuevos o significativamente mejorados.
6	Año_compra_software	1	Indica el año de compra de software y bases de datos.
7	Año_compra_afi	1	Indica el año de compra de otros activos fijos intangibles.

8	Energia_Alternativa	0.97	Indica si la empresa generó energía eléctrica alternativa o complementaria a la energía de la red pública.
9	Año_compra_marcas	0.9	Indica el año de compra de marcas.
10	Certificado_Digital	0.87	Indica si la página web de la empresa tenía certificado digital.
11	Empresa_MedianaB	0.86	Indica si la empresa es Mediana B.
12	Empresa_Grande	0.86	Indica si la empresa es Grande.
13	Año_RUC	0.86	Indica desde qué año la empresa ha tenido un RUC.
14	Edad	0.84	Indica la edad de la empresa
15	Sitio_Web	0.84	Indica si la empresa disponía de sitio o página web.
16	Credito_empleados	0.82	Indica si la empresa ofreció créditos a sus empleados.
17	Compras_por_Internet_nacional	0.8	Indica el porcentaje de ventas por internet que se realizó en el país.
18	PaginaInterna	0.78	Indica si la empresa contaba con una página interna en la Web (intranet/extranet)
19	MC_MediosImpresos	0.76	Indica si la empresa utilizó como medio de comunicación los medios impresos.
20	Cambio_tamano	0.75	Indica si la empresa cambió de tamaño en comparación al año anterior.
21	No_Cambio_tamano	0.75	Indica si la empresa no cambió de tamaño en comparación al año anterior.
22	Valor_elaboracion_CP_marcas	0.72	Indica el valor de elaboración por cuenta propia de Marcas.
23	Año_elaboracion_CP_afi	0.72	Indica el año de elaboración por cuenta propia de otros activos fijos intangibles.
24	Consumo_Combustible	0.71	Indica si la empresa consumió combustibles y/o lubricantes en diferentes procesos.

Tabla 35 - Predictores estables de la estrategia híbrida (con categóricas multinivel, partición = 10).

<u>No.</u>	<u>Código</u>	$\hat{\pi}_j$	<u>Descripción</u>
1	PA_Especialista_TIC	1	Indica si tiene personal afiliado especialista TIC.
2	Innovación	1	Indica si la empresa realizó actividades de investigación y desarrollo orientados a la generación de productos/procesos nuevos o significativamente mejorados.
3	Compras_por_Internet	1	Indica si la empresa compró bienes o servicios por internet.
4	Sitio_Web	1	Indica si la empresa disponía de sitio o página web.
5	Region_Costa	0.99	Indica si la empresa pertenece a la región Costa.
6	Region_Amazonia	0.99	Indica si la empresa pertenece a la región Amazonía.
7	Region_Insular	0.99	Indica si la empresa pertenece a la región Insular.
8	Edad	0.97	Indica la edad de la empresa
9	Total_PA_Mujeres	0.89	Indica el total de mujeres.

Tabla 36 - Predictores estables de la estrategia híbrida (con categóricas multinivel, partición = 5).

<u>No.</u>	<u>Código</u>	$\hat{\pi}_j$	<u>Descripción</u>
1	PA_Especialista_TIC	1	Indica si tiene personal afiliado especialista TIC.
2	Innovación	1	Indica si la empresa realizó actividades de investigación y desarrollo orientados a la generación de productos/procesos nuevos o significativamente mejorados.

3	Compras_por_Internet	1	Indica si la empresa compró bienes o servicios por internet.
4	Sitio_Web	1	Indica si la empresa disponía de sitio o página web.
5	Region_Costa	0.99	Indica si la empresa pertenece a la región Costa.
6	Region_Amazonia	0.99	Indica si la empresa pertenece a la región Amazonía.
7	Region_Insular	0.99	Indica si la empresa pertenece a la región Insular.
8	Edad	0.97	Indica la edad de la empresa
9	Total_PA_Mujeres	0.9	Indica el total de mujeres.

Tabla 37 - Predictores estables de la estrategia híbrida (sin categóricas multinivel, partición = 10).

<u>No.</u>	<u>Código</u>	$\hat{\pi}_j$	<u>Descripción</u>
1	PA_Especialista_TIC	1	Indica si tiene personal afiliado especialista TIC.
2	Innovación	1	Indica si la empresa realizó actividades de investigación y desarrollo orientados a la generación de productos/procesos nuevos o significativamente mejorados.
3	Compras_por_Internet	1	Indica si la empresa compró bienes o servicios por internet.
4	Sitio_Web	1	Indica si la empresa disponía de sitio o página web.
5	Internet	0.99	Indica si la empresa disponía de conexión a internet.
6	Total_PA_Mujeres	0.87	Indica el total de mujeres.
7	Edad	0.71	Indica la edad de la empresa

8	FirmaDigital	0.7	Indica si la empresa usó firma digital en comunicaciones enviadas.
---	--------------	-----	--

Tabla 38 - Predictores estables de la estrategia híbrida (sin categóricas multinivel, partición = 5).

<u>No.</u>	<u>Código</u>	$\hat{\pi}_j$	<u>Descripción</u>
1	PA_Especialista_TIC	1	Indica si tiene personal afiliado especialista TIC.
2	Innovación	1	Indica si la empresa realizó actividades de investigación y desarrollo orientados a la generación de productos/procesos nuevos o significativamente mejorados.
3	Compras_por_Internet	1	Indica si la empresa compró bienes o servicios por internet.
4	Sitio_Web	1	Indica si la empresa disponía de sitio o página web.
5	Internet	0.99	Indica si la empresa disponía de conexión a internet.
6	Total_PA_Mujeres	0.89	Indica el total de mujeres.
7	Edad	0.74	Indica la edad de la empresa
8	FirmaDigital	0.71	Indica si la empresa usó firma digital en comunicaciones enviadas.

Anexo D. Importancia Relativa Ponderada

Tabla 39 - Importancia ponderada de los predictores obtenidos mediante la estrategia de regularización (sin categóricas multinivel, partición = 10).

<u>No.</u>	<u>Código</u>	<u>Promedio</u> <u>ponderado</u>	<u>Desviación</u> <u>estándar</u>	<u>Frecuencia</u>
1	Salarios_Mujeres	14.93	5.80	7
2	Salarios_Nov_Mujeres	14.47	4.89	7
3	PA_ApoyoAd_Mujeres	14.10	3.22	7
4	Salarios_Nov_ApoyoAd_Mujeres	13.79	5.23	7
5	Innovación	11.36	5.08	7
6	Año_compra_software	8.56	4.61	7
7	PA_Especialista_TIC	6.04	5.04	6
8	Certificado_Digital	5.36	4.70	6
9	Sitio_Web	5.01	4.37	6
10	Uso_Internet_Capacitacion	4.76	3.93	6
11	Año_compra_añi	3.92	3.41	6
12	Compras_por_Internet	3.62	3.36	5
13	Uso_Internet_Contratacion	3.59	3.43	6
14	Año_RUC	3.55	3.33	6
15	PaginaInterna	3.45	3.63	5
16	MC_MediosImpresos	3.33	2.93	6
17	Uso_TIC_RRHH	3.28	2.71	6
18	Uso_TIC_Conocimiento	3.26	3.24	6
19	Compras_por_Internet_porcentaje	3.22	3.64	5
20	Organizaciones_Consultores	3.15	3.01	6
21	Edad	2.96	3.54	5
22	Red_LAN	2.95	2.52	6
23	MC_CorreoElectronico	2.86	2.79	6
24	Energia_Alternativa	2.48	1.95	7
25	Uso_Internet_Nube	2.45	2.12	6
26	Uso_Internet_Informacion	2.24	1.91	5
27	Interaccion_EG_ConseguirFormularios	1.99	1.28	6

28	Año_compra_marcas	1.99	1.95	6
29	Credito_empleados	1.96	1.83	6
30	Interaccion_EG_GestionElectronica	1.50	1.12	6
31	SoftwareAbierto_AplicacionesOfimaticas	1.21	1.22	4
32	Valor_elaboracion_CP_marcas	1.00	1.61	3

Ilustración 4 - Importancia ponderada de los 10 principales predictores obtenidos mediante la estrategia de regularización (sin categóricas multinivel, partición = 10).

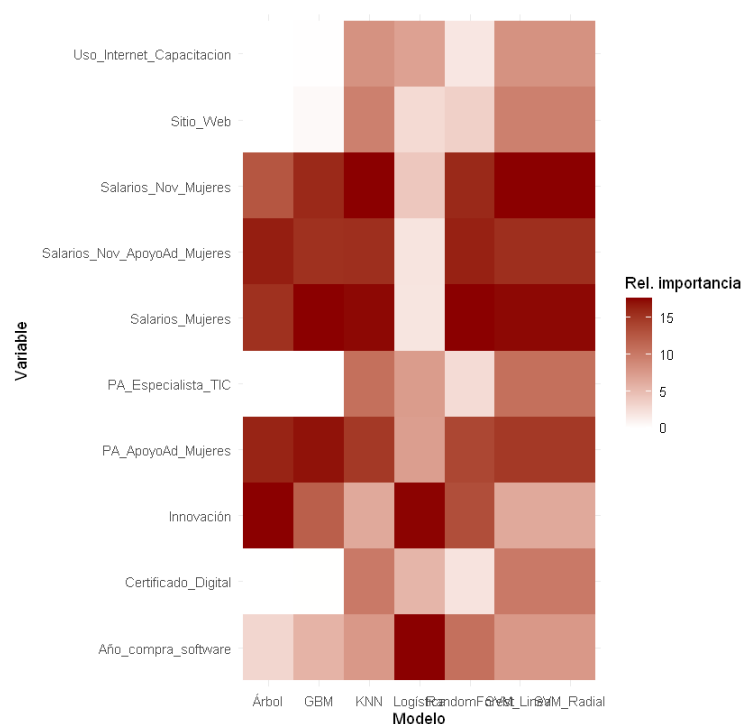


Tabla 40 - Importancia ponderada de los predictores obtenidos mediante la estrategia de regularización (sin categóricas multinivel, partición = 5).

<u>No.</u>	<u>Código</u>	<u>Promedio ponderado</u>	<u>Desviación estándar</u>	<u>Frecuencia</u>
1	Salarios_Nov_Mujeres	15.30	5.59	7
2	Salarios_Mujeres	14.99	5.50	7
3	PA_ApoyoAd_Mujeres	13.46	4.82	7
4	Salarios_Nov_ApoyoAd_Mujeres	12.71	5.09	7
5	Innovación	10.23	3.91	7

6	Año_compra_software	7.72	5.07	7
7	PA_Especialista_TIC	6.52	5.25	6
8	Sitio_Web	5.01	4.56	6
9	Certificado_Digital	4.90	4.98	6
10	Uso_Internet_Capacitacion	4.10	4.09	5
11	Compras_por_Internet_porcentaje	3.90	3.52	6
12	PaginaInterna	3.79	3.56	6
13	Organizaciones_Consultores	3.75	3.13	6
14	Uso_Internet_Contratacion	3.61	3.48	6
15	Red_LAN	3.55	2.94	6
16	Uso_TIC_Conocimiento	3.48	3.19	6
17	Uso_TIC_RRHH	3.37	2.76	6
18	Año_RUC	3.35	3.38	6
19	Año_compra_afi	3.15	2.47	6
20	Compras_por_Internet	3.09	3.65	6
21	Edad	3.04	3.57	5
22	MC_MediosImpresos	2.80	3.01	6
23	Credito_empleados	2.27	2.36	6
24	Energia_Alternativa	2.05	1.71	6
25	Uso_Internet_Informacion	1.95	1.65	6
26	Uso_Internet_Nube	1.93	2.14	6
27	MC_CorreoElectronico	1.85	1.15	6
28	Año_compra_marcas	1.84	1.51	7
29	Interaccion_EG_ConseguirFormularios	1.70	1.06	6
30	Interaccion_EG_GestionElectronica	1.61	1.09	6
31	SoftwareAbierto_AplicacionesOfimaticas	1.33	1.43	5
32	Valor_elaboracion_CP_marcas	1.04	1.60	3

Ilustración 5 - Importancia ponderada de los principales predictores obtenidos mediante la estrategia de regularización (sin categóricas multinivel, partición = 5).

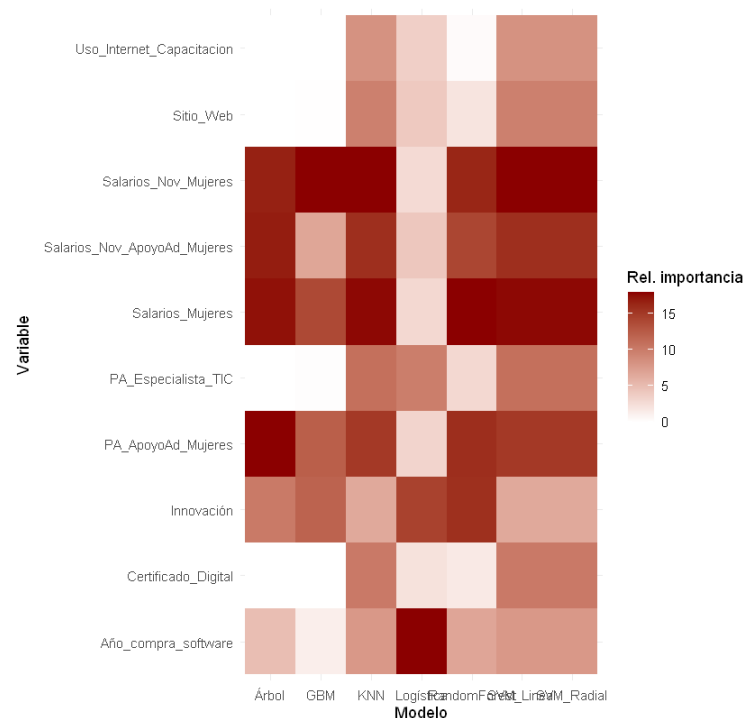


Tabla 41 - Importancia ponderada de los predictores obtenidos mediante la estrategia de regularización (con categóricas multinivel, partición = 10).

<u>No.</u>	<u>Código</u>	<u>Promedio</u> <u>ponderado</u>	<u>Desviación</u> <u>estándar</u>	<u>Frecuencia</u>
1	Adq_Muebles	5.59	6.24	5
2	Endeudamiento_Patrimonio	5.37	6.39	7
3	Valor_elaboracion_CP_marcas	5.29	5.90	5
4	PA_ApoyoAd_Mujeres	5.17	5.59	4
5	Salarios_Nov_Directores_Hombres	4.56	5.48	5
6	PA_Especialista_TIC	4.39	4.13	5
7	Gastos_seguros	4.09	4.19	5
8	Dispositivos_PDA	4.05	4.79	5
9	Certificado_Digital	3.99	3.79	5
10	Año_compra_software	3.80	3.80	5
11	Innovación	3.78	4.41	4
12	Adq_AFU_Maquinaria	3.56	3.98	7
13	Sitio_Web	3.41	3.56	4
14	Uso_Internet_Capacitacion	3.33	3.13	5

15	Duracion_AE_CP_derechosLlave	3.19	3.56	7
16	Gastos_indirectos_exterior	3.16	3.47	4
17	Costo_PromocionPublicidad	2.97	3.60	5
18	Uso_Internet_Contratacion	2.82	2.68	5
19	Año_elaboracion_CP_software	2.81	2.94	5
20	Compras_por_Internet_porcentaje	2.65	2.64	5
21	Compras_por_Internet_nacional	2.54	2.54	4
22	PaginaInterna	2.54	2.74	5
23	Empresa_Grande	2.49	2.46	5
24	MC_MediosImpresos	2.42	2.27	4
25	Año_compra_afi	2.39	2.24	4
26	Utilidad_venta_AF_Maquinaria	2.36	2.55	4
27	Cambio_tamano	2.32	2.22	5
28	Organizaciones_Consultores	2.29	2.36	5
29	Uso_TIC_Conocimiento	2.27	2.52	5
30	Uso_TIC_RRHH	2.25	2.15	4
31	Duracion_AC_AFI	2.22	2.11	4
32	Construcciones_AF_CP_OtrosAFI	2.21	2.63	4
33	Adq_AFU_EquipoCamionero	2.20	2.95	7
34	ID_Externa	2.19	1.77	6
35	Red_LAN	2.16	2.07	4
36	Region_Costa	2.14	2.01	4
37	Adq_AFU_EquipoComputacion	2.04	2.33	6
38	Edad	2.04	2.48	5
39	CS_nacional_publico	1.97	2.04	4
40	Compras_por_Internet_extranjero	1.94	2.38	5
41	Uso_Internet_Nube	1.79	1.68	5
42	Año_elaboración_CP_derechoLlave	1.78	1.96	5
43	Uso_Internet_Informacion	1.65	1.58	4
44	Credito_empleados	1.63	1.69	4
45	Duracion_AE_CP_marcas	1.62	1.53	4
46	Consumo_Gasolina_galones	1.62	1.53	4
47	Region_Insular	1.30	1.86	6

48	Empresa_MedianaB	1.27	1.22	5
49	Salarios_Nov_Operadores_Mujeres	1.21	1.16	4
50	Cantidad_Agua_Consumida	1.07	1.29	4
51	Año_RUC	1.04	1.53	7
52	Venta_AF_Maquinarias	0.97	1.09	4
53	Const_AF_CP_Edificios	0.91	1.30	5
54	CS_nacional_privado	0.89	0.90	5
55	Energia_Alternativa	0.83	0.81	5
56	Año_compra_marcas	0.81	0.90	4
57	Costos_impuestos	0.78	0.73	5
58	Costos_ajustes	0.74	0.79	5
59	Consumo_Combustible	0.70	0.66	5
60	Año_elaboracion_CP_afi	0.70	1.52	4
61	Region_Amazonia	0.61	1.05	6
62	Ganancias_netas_actividades_discontinuas	0.60	0.62	4
63	No_Cambio_tamano	0.60	0.58	5
64	Costo_viaje	0.55	1.45	5
65	Regalias_cesiones	0.53	1.35	4
66	Obst_VI_Caracteristicas	0.49	1.08	5
67	Año_compra_patentes	0.44	0.97	5
68	Valor_GasolinaSuper	0.43	0.41	4
69	Credito_otros	0.37	0.51	4
70	Credito_banco_privado	0.36	0.90	5
71	Salarios_Nov_ApoyoAd_Mujeres	0.36	0.93	5
72	Costo_transporte	0.35	0.34	4
73	Costos_gastos_Operaciones	0.33	0.60	5
74	Duracion_AC_franquicia	0.30	0.65	4
75	Año_compra_derechoLlave	0.27	0.35	5
76	Dep_EquipoCamionero	0.27	0.71	5
77	Adq_Vehiculos	0.26	0.33	5
78	Venta_AF_edificios	0.25	0.34	5
79	Revalorizaciones_Terrenos	0.24	0.56	5

80	Intereses_IF	0.23	0.57	4
81	Utilidad_venta_AF_Terrenos	0.20	0.25	4
82	Valor_produccion_Energia_Solar	0.18	0.24	4
83	Costo_Honorarios	0.17	0.44	1
84	Salarios_Nov_TNC	0.16	0.41	4
85	CS_extranjero_publico	0.12	0.32	4
86	Adq_AFN_Vehiculos	0.12	0.15	5
87	Venta_energia_BiomasaProducida	0.08	0.07	5
88	Utilidad_venta_AF_Abiologicos	0.05	0.11	5

Ilustración 6 - Importancia ponderada de los principales predictores obtenidos mediante la estrategia de regularización (con categóricas multinivel, partición = 10).

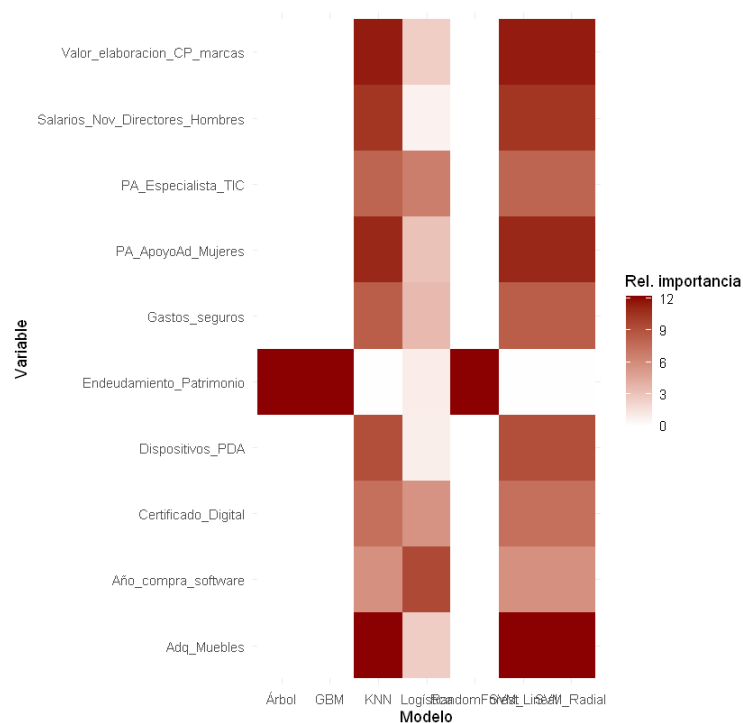


Tabla 42 - Importancia ponderada de los predictores obtenidos mediante la estrategia de regularización (con categóricas multinivel, partición = 5).

<u>No.</u>	<u>Código</u>	<u>Promedio</u> <u>ponderado</u>	<u>Desviación</u> <u>estándar</u>	<u>Frecuencia</u>
1	Adq_AFN_Muebles	6.20	4.77	5
2	Adq_Muebles	5.94	4.82	5
3	PA_ApoyoAd_Mujeres	5.07	4.32	5
4	Salarios_Nov_ApoyoAd_Mujeres	5.05	4.66	6
5	Salarios_Nov_Directores_Hombres	4.73	4.20	5
6	Gastos_seguros	4.35	3.43	6
7	Adq_AFU_Maquinaria	4.04	4.87	7
8	Año_compra_software	3.86	3.63	5
9	PA_Especialista_TIC	3.86	3.37	5
10	Innovación	3.72	3.21	6
11	Adq_AFN_Vehiculos	3.69	2.76	5
12	Adq_Vehiculos	3.50	2.70	5
13	Dispositivos_PDA	3.44	4.09	5
14	Año_elaboración_CP_derechoLlave	3.27	4.12	6
15	Certificado_Digital	3.22	3.08	5
16	Sitio_Web	3.04	2.91	5
17	Cantidad_Agua_Consumida	3.03	2.34	6
18	Duracion_AE_CP_derechosLlave	2.81	4.45	6
19	Adq_AFU_EquipoComputacion	2.68	3.40	7
20	Uso_Internet_Capacitacion	2.62	2.56	5
21	Uso_Internet_Contratacion	2.44	2.25	6
22	PaginaInterna	2.34	2.27	6
23	Red_LAN	2.33	2.22	5
24	Organizaciones_Consultores	2.28	2.04	6
25	Adq_AFU_EquipoCamionero	2.22	3.51	7
26	Duracion_AC_AFI	2.21	1.82	5
27	Salarios_Nov_TNC	2.07	1.96	5
28	Empresa_Grande	2.02	1.78	5
29	Compras_por_Internet_porcentaje	1.97	2.38	5

30	Compras_por_Internet_extranjero	1.97	1.90	5
31	Edad	1.92	2.21	6
32	Año_RUC	1.88	2.26	5
33	MC_MediosImpresos	1.87	1.85	5
34	Uso_TIC_RRHH	1.86	1.75	5
35	Uso_TIC_Conocimiento	1.85	2.21	5
36	Region_Costa	1.74	1.59	5
37	Cambio_tamano	1.74	1.90	5
38	Compras_por_Internet_nacional	1.71	1.64	5
39	Año_compra_afi	1.61	1.75	5
40	Costos_ajustes	1.46	2.12	5
41	Consumo_Gasolina_galones	1.31	1.26	5
42	Salarios_Nov_Operadores_Mujeres	1.27	1.26	6
43	Venta_AF_edificios	1.26	2.64	6
44	Credito_empleados	1.25	1.19	6
45	Const_AF_CP_Edificios	1.22	1.97	5
46	Uso_Internet_Informacion	1.20	1.08	5
47	Endeudamiento_Patrimonio	1.09	2.30	6
48	Empresa_MedianaB	1.08	0.98	5
49	Energia_Alternativa	1.06	0.97	5
50	CS_nacional_privado	1.03	1.03	5
51	Credito_banco_privado	1.03	1.34	6
52	Venta_AF_Maquinarias	0.83	0.94	5
53	Costo_PromocionPublicidad	0.83	1.85	5
54	Año_compra_marcas	0.82	0.77	5
55	Año_elaboracion_CP_software	0.78	1.21	6
56	Costos_impuestos	0.69	0.50	5
57	Consumo_Combustible	0.60	0.56	5
58	Utilidad_venta_AF_Maquinaria	0.57	0.64	5
59	No_Cambio_tamano	0.51	0.47	5
60	Año_compra_patentes	0.51	0.82	5
61	Dep_EquipoCamionero	0.50	0.81	6
62	Costo_viaje	0.48	0.82	5

63	Valor_elaboracion_CP_marcas	0.48	0.94	5
64	Año_elaboracion_CP_afí	0.48	0.67	6
65	CS_nacional_publico	0.48	0.51	5
66	Revalorizaciones_Terrenos	0.47	0.83	5
67	Valor_produccion_Energia_Solar	0.46	0.77	5
68	Año_compra_derechoLlave	0.40	0.52	5
69	Regalias_cesiones	0.39	0.73	5
70	Credito_otros	0.38	0.56	5
71	Region_Insular	0.37	0.97	2
72	Gastos_indirectos_exterior	0.35	0.37	6
73	Venta_energia_BiomasaProducida	0.32	0.70	5
74	Region_Amazonia	0.23	0.37	5
75	Utilidad_venta_AF_Terrenos	0.23	0.35	5
76	Ganancias_netas_actividades_discontinuas	0.14	0.23	6
77	CS_extranjero_publico	0.13	0.33	5

Ilustración 7 - Importancia ponderada de los principales predictores obtenidos mediante la estrategia de regularización (con categóricas multinivel, partición = 5).

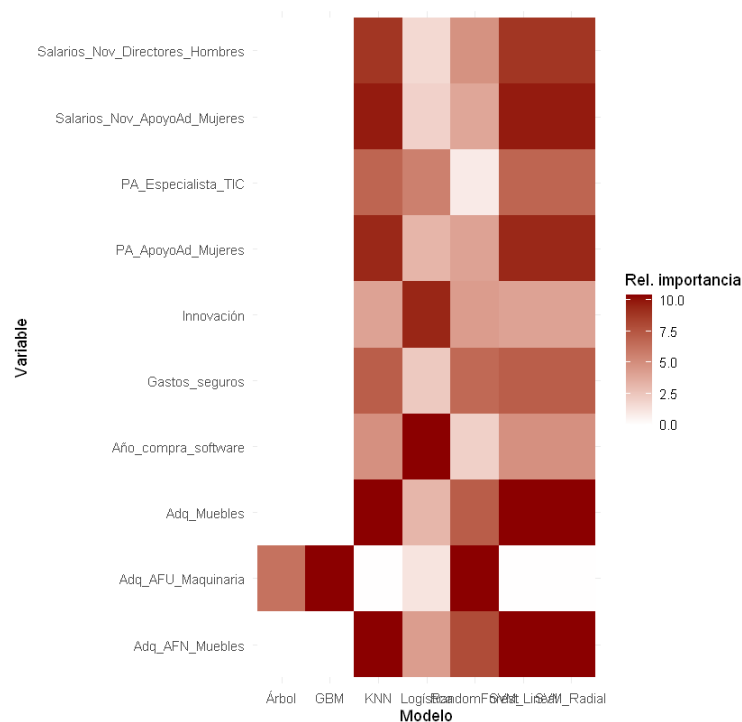


Tabla 43 - Importancia ponderada de los predictores obtenidos mediante la estrategia manual (con categóricas multinivel, partición = 10).

<u>No.</u>	<u>Código</u>	<u>Promedio ponderado</u>	<u>Desviación estándar</u>	<u>Frecuencia</u>
1	PA_Especialista_TIC	2.09	1.97	5
2	Sitio_Web	1.87	1.75	5
3	Innovación	1.66	1.31	6
4	ID_Externa	1.46	1.72	7
5	Internet	1.36	1.61	7
6	Compras_por_Internet	1.16	1.10	5
7	Total_PA_Mujeres	1.09	1.01	5
8	Edad	1.00	0.95	5
9	Region_Costa	0.94	0.88	5
10	Total_PA	0.92	0.91	5
11	PA_Científicos	0.78	0.74	5
12	VAE	0.78	0.77	5
13	Endeudamiento_Patrimonio	0.77	0.88	7
14	Ventas	0.70	0.67	5
15	ID_Interna	0.68	0.72	7
16	ID_Departamento	0.57	0.56	5
17	Ventas_por_Internet	0.56	0.67	5
18	Sector_Comercio	0.56	0.67	5
19	Sector_Servicios	0.27	0.28	4
20	FirmaDigital	0.26	0.27	5
21	Sector_Mineria	0.12	0.15	4
22	Region_Amazonia	0.10	0.24	5
23	Region_Insular	0.09	0.16	5
24	Sector_Construccion	0.00	0.01	2

Ilustración 8 - Importancia ponderada de los principales predictores obtenidos mediante la estrategia manual (con categóricas multinivel, partición = 10).

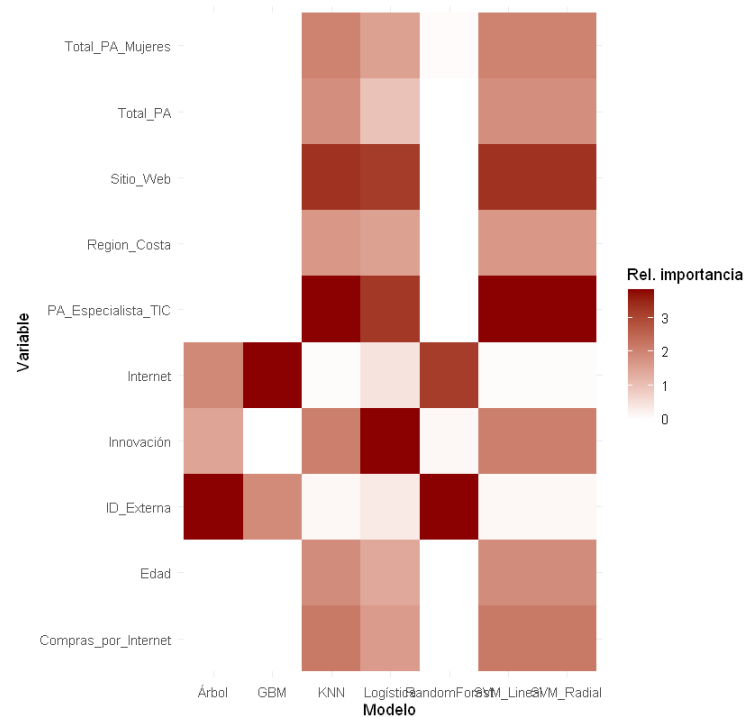


Tabla 44 - Importancia ponderada de los predictores obtenidos mediante la estrategia manual (con categóricas multinivel, partición = 5).

<u>No.</u>	<u>Código</u>	<u>Promedio ponderado</u>	<u>Desviación estándar</u>	<u>Frecuencia</u>
1	PA_Especialista_TIC	2.09	1.97	4
2	Sitio_Web	1.87	1.75	5
3	ID_Externa	1.73	1.94	7
4	Innovación	1.66	1.31	6
5	Compras_por_Internet	1.15	1.10	5
6	Endeudamiento_Patrimonio	1.14	1.14	7
7	Total_PA_Mujeres	1.08	1.01	5
8	Internet	1.04	1.14	7
9	Edad	1.00	0.95	5
10	Region_Costa	0.94	0.88	5
11	Total_PA	0.92	0.91	5
12	PA_Cientificos	0.78	0.74	5

13	VAE	0.78	0.77	5
14	Ventas	0.71	0.67	5
15	ID_Interna	0.69	0.67	7
16	ID_Departamento	0.57	0.56	5
17	Ventas_por_Internet	0.56	0.67	5
18	Sector_Comercio	0.56	0.67	5
19	Sector_Servicios	0.27	0.28	5
20	FirmaDigital	0.26	0.27	5
21	Sector_Mineria	0.12	0.15	4
22	Region_Amazonia	0.10	0.24	5
23	Region_Insular	0.09	0.16	5
24	Sector_Construccion	0.00	0.01	2

Ilustración 9 - Importancia ponderada de los principales predictores obtenidos mediante la estrategia manual (con categóricas multinivel, partición = 5).



Tabla 45 - Importancia ponderada de los predictores obtenidos mediante la estrategia manual (sin categóricas multinivel, partición = 10).

<u>No.</u>	<u>Código</u>	<u>Promedio ponderado</u>	<u>Desviación estándar</u>	<u>Frecuencia</u>
1	PA_Especialista_TIC	1.84	1.66	5
2	Sitio_Web	1.53	1.37	5
3	Innovación	1.48	0.95	6
4	ID_Externa	1.44	1.65	7
5	Internet	1.12	1.33	7
6	Total_PA_Mujeres	1.02	0.72	5
7	Compras_por_Internet	0.96	0.90	5
8	PA_Científicos	0.84	0.58	5
9	ID_Interna	0.82	1.08	7
10	Total_PA	0.81	0.66	5
11	Endeudamiento_Patrimonio	0.81	1.11	4
12	Ventas	0.74	0.52	5
13	Edad	0.67	0.65	5
14	VAE	0.66	0.53	5
15	Ventas_por_Internet	0.50	0.58	4
16	ID_Departamento	0.43	0.42	5
17	FirmaDigital	0.14	0.16	4

Ilustración 10 - Importancia ponderada de los principales predictores obtenidos mediante la estrategia manual (sin categóricas multinivel, partición = 10).

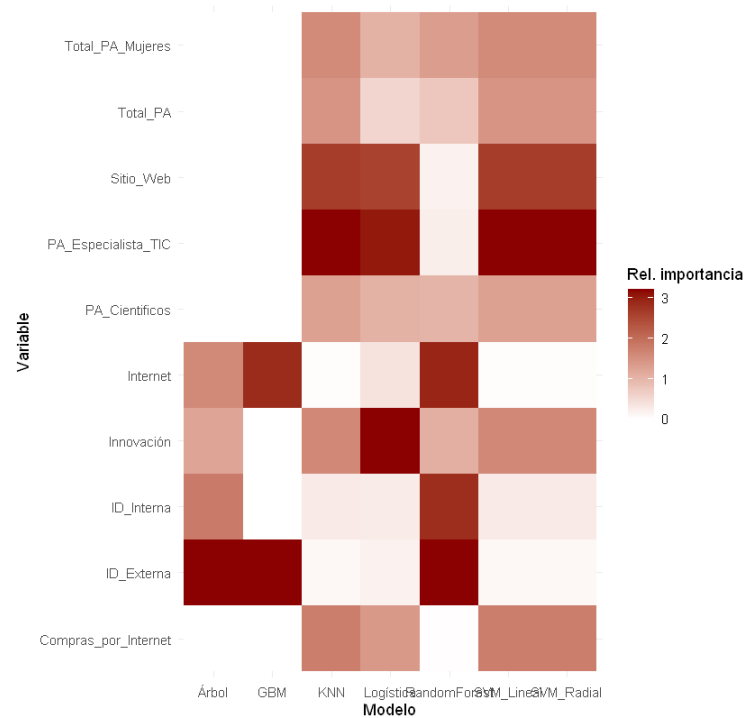


Tabla 46 - Importancia ponderada de los predictores obtenidos mediante la estrategia manual (sin categóricas multinivel, partición = 5).

<u>No.</u>	<u>Código</u>	<u>Promedio ponderado</u>	<u>Desviación estándar</u>	<u>Frecuencia</u>
1	PA_Especialista_TIC	1.93	1.75	5
2	Sitio_Web	1.61	1.45	5
3	Innovación	1.56	1.00	6
4	ID_Externa	1.34	1.57	7
5	Internet	1.17	1.40	7
6	Total_PA_Mujeres	1.04	0.75	5
7	Compras_por_Internet	1.00	0.94	5
8	Endeudamiento_Patrimonio	0.89	1.10	4
9	PA_Cientificos	0.88	0.61	5
10	ID_Interna	0.83	1.05	7
11	Total_PA	0.83	0.70	5

12	Ventas	0.78	0.55	5
13	Edad	0.70	0.69	5
14	VAE	0.67	0.56	5
15	Ventas_por_Internet	0.52	0.61	4
16	ID_Departamento	0.45	0.45	5
17	FirmaDigital	0.14	0.16	4

Ilustración 11 - Importancia ponderada de los principales predictores obtenidos mediante la estrategia manual (sin categóricas multinivel, partición = 5).



Tabla 47 - Importancia ponderada de los predictores obtenidos mediante la estrategia híbrida (con categóricas multinivel, partición = 10).

<u>No.</u>	<u>Código</u>	<u>Promedio</u> <u>ponderado</u>	<u>Desviación</u> <u>estándar</u>	<u>Frecuencia</u>
1	Total_PA_Mujeres	4.14	3.01	6
2	Total_Dispositivos	3.91	4.01	5
3	PA_UsaInternet_Mujeres	3.83	3.41	5

4	Total_PA	3.28	3.21	5
5	VAE	3.25	3.38	5
6	ROE	3.15	3.87	7
7	Innovación	3.10	2.45	6
8	Endeudamiento_Patrimonio	3.00	3.80	6
9	PA_Especialista_TIC	2.97	2.63	5
10	ID_Externa	2.74	2.85	7
11	Sitio_Web	2.64	2.31	5
12	Compras_por_Internet	1.79	1.62	5
13	Edad	1.55	1.49	5
14	Region_Costa	1.36	1.27	5
15	FirmaDigital	0.74	0.75	5
16	Region_Amazonia	0.08	0.06	5
17	Region_Insular	0.07	0.18	1

Ilustración 12 - Importancia ponderada de los principales predictores obtenidos mediante la estrategia híbrida (con categóricas multinivel, partición = 10).



Tabla 48 - Importancia ponderada de los predictores obtenidos mediante la estrategia híbrida (con categóricas multinivel, partición = 5).

<u>No.</u>	<u>Código</u>	<u>Promedio ponderado</u>	<u>Desviación estándar</u>	<u>Frecuencia</u>
1	PA_UsaInternet_Mujeres	5.25	3.49	7
2	Total_Dispositivos	4.57	4.85	6
3	Total_PA_Mujeres	4.47	4.00	6
4	Endeudamiento_Patrimonio	4.09	5.05	6
5	VAE	3.93	4.00	6
6	Total_PA	3.76	3.97	6
7	Innovación	3.75	2.91	7
8	PA_Especialista_TIC	3.74	3.44	6
9	ID_Externa	3.32	3.40	7
10	Sitio_Web	3.13	2.77	6
11	ROE	3.11	3.86	7
12	Compras_por_Internet	2.00	1.85	6
13	Edad	1.86	1.77	6
14	Region_Costa	1.49	1.35	6
15	Region_Amazonia	0.22	0.37	6
16	Region_Insular	0.22	0.58	1

Ilustración 13 - Importancia ponderada de los principales predictores obtenidos mediante la estrategia híbrida (con categóricas multinivel, partición = 5).

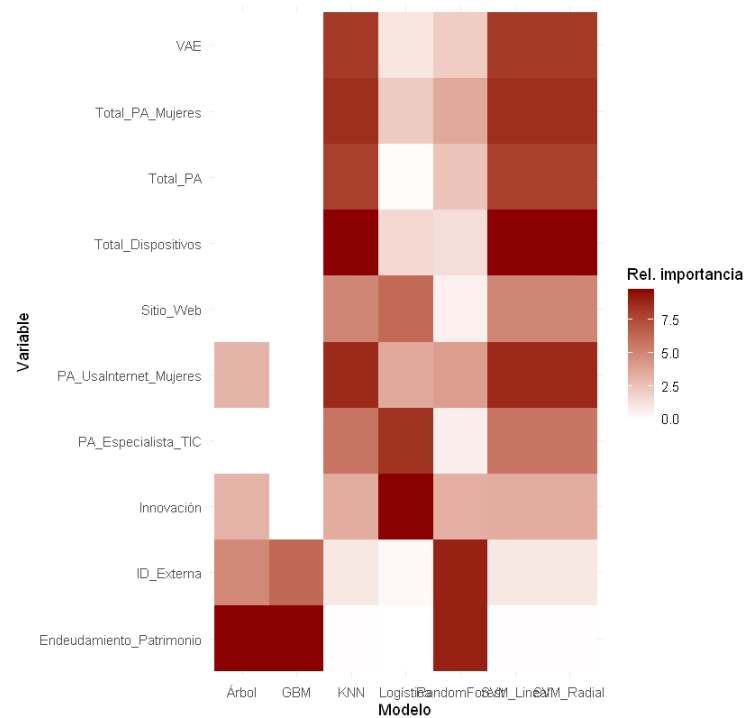


Tabla 49 - Importancia ponderada de los predictores obtenidos mediante la estrategia híbrida (sin categóricas multinivel, partición = 10).

<u>No.</u>	<u>Código</u>	<u>Promedio</u> <u>ponderado</u>	<u>Desviación</u> <u>estándar</u>	<u>Frecuencia</u>
1	Total_PA_Mujeres	4.31	3.04	6
2	Total_Dispositivos	4.17	4.02	6
3	Internet	3.61	4.51	3
4	VAE	3.58	3.30	6
5	Innovación	3.12	2.55	6
6	PA_Especialista_TIC	2.95	2.65	5
7	Sitio_Web	2.62	2.27	5
8	Compras_por_Internet	1.85	1.71	5
9	Edad	1.35	1.54	4
10	FirmaDigital	0.53	0.48	5

Ilustración 14 - Importancia ponderada de los principales predictores obtenidos mediante la estrategia híbrida (sin categóricas multinivel, partición = 10).

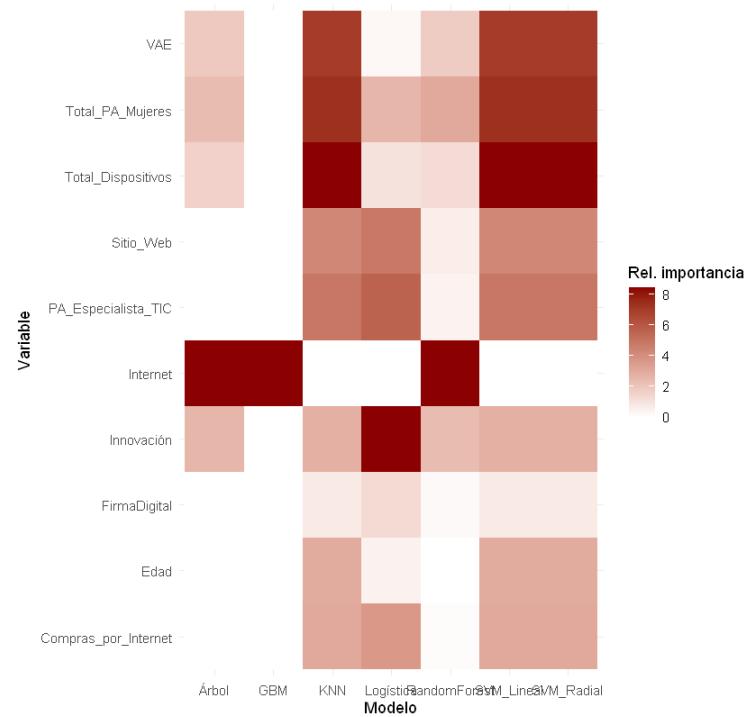


Tabla 50 - Importancia ponderada de los predictores obtenidos mediante la estrategia híbrida (sin categóricas multinivel, partición = 5).

<u>No.</u>	<u>Código</u>	<u>Promedio ponderado</u>	<u>Desviación estándar</u>	<u>Frecuencia</u>
1	Total_PA_Mujeres	4.15	2.87	6
2	Total_Dispositivos	3.79	3.98	5
3	VAE	3.42	3.11	6
4	Internet	3.18	3.68	7
5	PA_Especialista_TIC	3.10	2.89	5
6	Innovación	3.02	2.41	6
7	Endeudamiento_Patrimonio	2.97	3.86	3
8	Sitio_Web	2.43	2.15	5
9	Compras_por_Internet	1.78	1.66	5
10	Edad	1.32	1.50	4
11	FirmaDigital	0.43	0.42	5

Ilustración 15 - Importancia ponderada de los principales predictores obtenidos mediante la estrategia híbrida (sin categóricas multinivel, partición = 5).

