LA GESTIÓN DEL RIESGO DE CRÉDITO EN LA ERA DE LA INTELIGENCIA ARTIFICIAL

IGNACIO JIMÉNEZ

LUIS HERNÁNDEZ

IBERINFORM

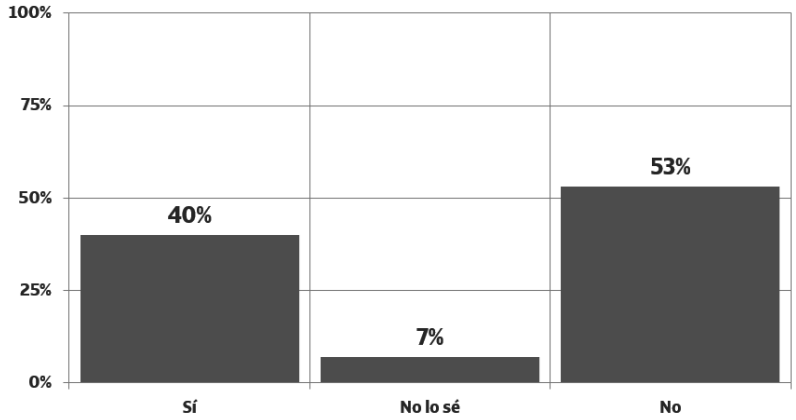
Cuando la economía va bien, son muchas las empresas que se olvidan de que vender no es lo mismo que cobrar, y que cobrar tampoco es lo mismo que disponer del dinero en la cuenta. Normalmente, el deterioro de la capacidad de pago de una empresa o autónomo se produce de forma gradual, pero las señales de alerta que identifican ese riesgo pueden surgir de múltiples fuentes de información y además no son fáciles de interpretar, lo que complica su gestión.

Tener poca información genera importantes riesgos, pero tener demasiada y mal estructurada sólo lleva a la «infoxicación» (intoxicación de información) y puede producir riegos igualmente importantes si bloquea la toma de decisiones. El incumplimiento de los pagos acordados con los clientes genera importantes tensiones de liquidez en las empresas, pero es especialmente desestabilizadora para aquellas estructuras de menor tamaño.

La necesidad de gestionar esa ingente cantidad de información ha llevado a las empresas de mayor tamaño a desarrollar áreas para gestionar la información y sus riesgos empresariales de una forma profesional, objetiva y eficiente. Si se produce el impago de una venta a crédito, la pérdida suele superar los costes de producción del producto o de prestar el servicio y su impacto se acrecienta significativamente cuanto menor es el margen de beneficio con el que opera la empresa, pues multiplica el número de ventas con clientes solventes que es necesario lograr para compensar dicha pérdida. Si una empresa tiene un margen comercial del 10% y sufre un impago de 10.000 euros, deberá generar nuevo negocio equivalente a 100.000 euros para compensar el impacto de los 9.000 en costes de producción.

Las empresas de mayor tamaño son conscientes de la importancia de gestionar este riesgo asociado a las ventas a crédito y cuentan con sistemas, protocolos y equipos que les permiten minimizar ese tipo de riesgos. En el caso de las entidades financieras, la gestión de riesgo de crédito forma parte de su negocio principal, por lo que son las que mayor cantidad de recursos dedican a fortalecer esa capacidad de anticipar y gestionar esos riesgos, y han sido también las que mejor han sabido aprovechar las nuevas tecnologías para desarrollar dichas capacidades. Igualmente fueron de las primeras empresas que invirtieron en informatizar los procesos de análisis de riesgo, llevan años utilizado las tecnologías asociadas al Big Data (BD) para recopilar, analizar y accionar la inmensa cantidad de datos que manejan, han construido potentes algoritmos para simplificar

FIGURA 1
DETERIORO DE SOLVENCIA POR EL COVID-19



Fuente: Iberinform y Crédito y Caución. Estudio de Gestión del Riesgo de Crédito en España. Marzo 2021

el análisis de riesgos y es uno de los sectores que más está utilizando la inteligencia artificial (IA). Han generado unas capacidades que están utilizado también para mejorar sus procesos de captación y fidelización de clientes, ajustar sus ofertas o el precio de sus servicios, sus previsiones y, sobre todo, para mejorar la capacidad predictiva de sus modelos de gestión del riesgo y para evitar fraudes.

En las empresas de menor tamaño, por el contrario, la escasez de recursos hace que sean mucho más vulnerables al impacto de la morosidad. Debido a ello este análisis se va a centrar en la pequeña y mediana empresa (Pyme), que, si aplicamos el criterio de la Unión Europea (1), representa el 95% del tejido empresarial de España, siendo el segmento que más peso tiene en la economía.

Los gestores de las pymes suelen tener un trato más directo con los clientes que en aquellas empresas de una mayor dimensión y suelen ser muy cautelosos con los clientes nuevos, pero esa cercanía a menudo también genera sesgos en la percepción del riesgo comercial o de impago y les impide ver los signos de deterioro que llevan a una empresa o un autónomo a incumplir sus compromisos de pago. De hecho, si se le pregunta al responsable de cualquier empresa qué nivel de información tiene de sus clientes, seguro que afirmará que mantiene una excelente relación personal con los más importantes, que a menudo se reúne con ellos para comer, que tiene los datos de sus clientes en el CRM de la empresa o que en su contabilidad puede ver cuánto le facturan.

En una situación tan cambiante como la actual, tomar decisiones basadas en datos es cada vez más difícil y, a la vez, más importante. El problema es que muchas Pymes basan su gestión del riesgo de crédito en percepciones superficiales sobre datos internos inconexos o elementos externos, como pueden ser el aspecto de la oficina que tiene el cliente o el vehículo que conduce el propietario, sin un análisis objetivo ni criterios técnicos. De forma que es poco probable que sepan si un cliente les está comprando más porque otros proveedores no le venden por retrasos en sus pagos o si lleva meses sin poder pagar las cuotas del contrato de renting del vehículo que conduce o el alquiler de su oficina. Además, como hemos visto estos últimos años, el entorno puede cambiar de forma brusca y la capacidad de las empresas para analizar y gestionar el riesgo de impago de clientes B2B con efectividad y anticipación se convierte un elemento crítico para su supervivencia, sobre todo en aquellas con menor fortaleza financiera.

De hecho, según el *Estudio de Gestión del Riesgo de Crédito* (2) publicado en noviembre 2021 por Iberinform y Crédito y Caución, el 9% de las empresas, cerca de 127.000 de acuerdo con los datos de empresas activas en la Seguridad Social, afirma que corre el riesgo de cerrar por el impacto de los impagados. Con la peor parte de la crisis generada por la Covid-19 ya superada, el 40% de las empresas españolas también confirma que ha detectado un deterioro importante de los niveles de solvencia de sus clientes. Son datos que dan una idea del tremendo impacto que tiene la morosidad en el tejido productivo del país y la capacidad que tiene de afectar a empresas solventes.

Las distintas medias económicas que se tomaron en Europa, y la rapidez con la que se aplicaron, han suavizado el impacto de la crisis y han logrado que la morosidad no haya vuelto a los niveles que experimentamos en crisis la económica de 2008, pero cuando se retiren esos apoyos financieros los niveles de morosidad podrían escalar con rapidez y una parte importe de las pymes no estaría preparada para gestionar la situación.

En este contexto, la experiencia indica que la opción más habitual entre las empresas de menor tamaño

para conocer la salud financiera de los clientes sigue siendo preguntar a otros actores del sector, buscar información en internet o consultar manualmente la información disponible en alguna plataforma online. Estas medias suelen ser reactivas y pueden ayudar a mitigar el riesgo de impago, pero generan un consumo de recursos desproporcionado y tienen un alcance y efectividad limitada. Además, una gestión manual requiere de personal con experiencia en este ámbito, un recurso del que no dispone la mayor parte de las empresas, y además tampoco sería factible actualizar los análisis de forma recurrente.

En el caso que una pyme decida abordar un proyecto para mejorar la información de clientes que utiliza y su gestión del riesgo comercial, en primer lugar, debe identificar qué fuentes de información tiene a su disposición. Así que, a continuación, vamos a analizar las principales fuentes de información a las que podemos acudir para identificar aspectos como el nivel de riesgo que implica trabajar con un determinado cliente, con qué clientes tenemos más oportunidades de desarrollar negocio o cómo podemos generar señales de alerta que nos permitan gestionar los riesgos de forma proactiva. La información podemos agruparla en cuatro grandes grupos:

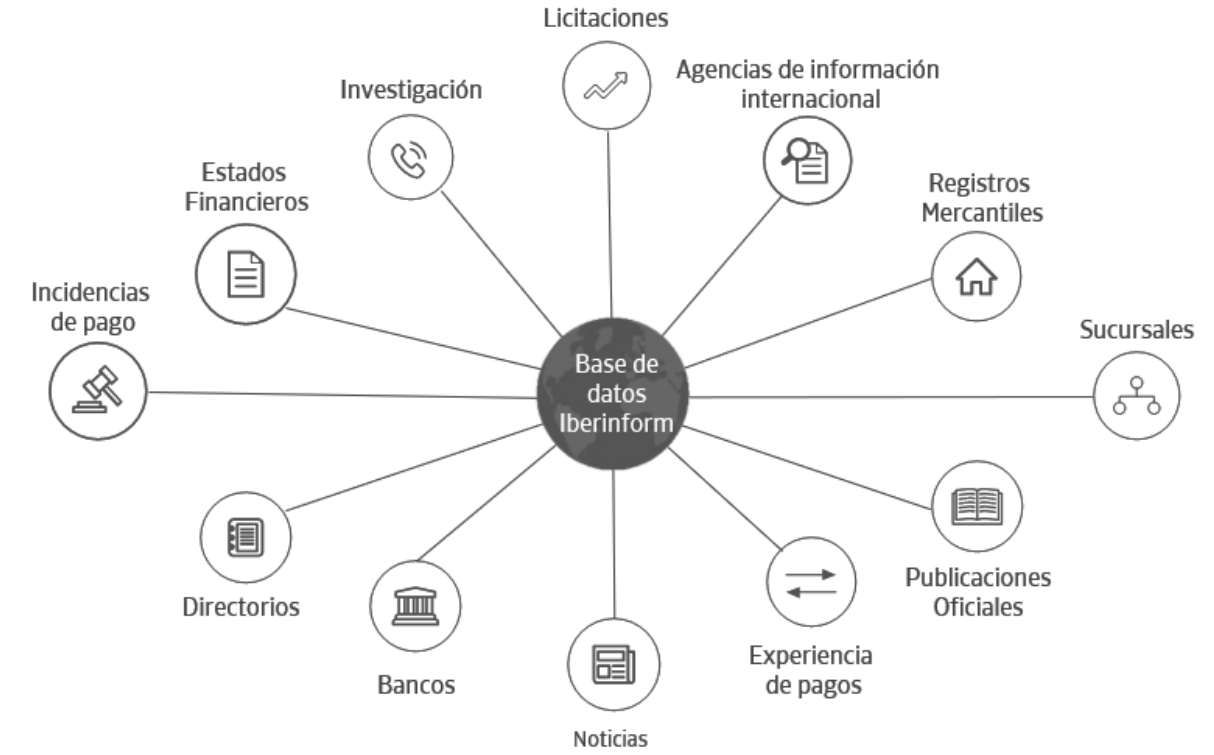
- La información interna que las empresas generan a través de la relación con sus clientes es muy rica, no tiene coste y, normalmente, está muy actualizada. Debería de ser el punto de partida de cualquier iniciativa que busque conocer en profundidad a sus clientes o valorar su riesgo comercial, pero cuando se deciden a trabajar con ella, la mayoría descubre que es mucho más complicado de lo que esperaban. La información está en silos y dispersa entre las diferentes áreas que la gestionan, normalmente carece de una estructura adecuada o simplemente no existe, y exige de grandes esfuerzos para limpiarla y normalizarla. Además, en algunos casos, puede estar sesgada por la relación personal que las propias Pymes tienen con sus clientes históricos o de mayor tamaño. Se pueden producir sesgos cognitivos a la hora de valorar el riesgo, es decir, aquellos «atajos» intuitivos y emocionales que nos pueden jugar una mala pasada a la hora de tomar decisiones. Es una información de gran valor, pero el coste que requiere exprimir su potencial demanda que las empresas tengan o contraten perfiles con conocimientos avanzados en la ingesta y análisis de datos, y esto suele ser una barrera infranqueable para la mayoría de las Pymes.

- Información informal obtenida de terceros, como por ejemplo otros proveedores del mismo sector o entidades financieras. Es una información de gran valor, pero dada las características de esta fuente es prácticamente imposible crear un modelo matemático para explotar estos datos, y tampoco podemos olvidar que también puede estar sesgada por los intereses del interlocutor y por los sesgos cognitivos de la persona que interpreta esa información. En cualquier caso, podemos valorar utilizar la información para crear reglas expertas que maticen el resultado del modelo, como por ejemplo en aquellos casos que tengamos información creíble de que el cliente no va a poder o querer atender a sus compromisos de pago.

- La información de fuentes como los registros oficiales, prensa, internet, redes sociales, etc. puede ser de libre acceso (*open data*) o estar en repositorios como el Registro de Mercantil o de la Propiedad, donde el acceso sí tiene coste. Este tipo de información es muy rica y objetiva, pero debe capturarse de fuentes enormemente dispersas, en formatos muy diversos (imágenes, textos, tablas, etc.) que sufre continuos cambios en la forma de acceder a los datos o en los propios contenidos, por lo que entender algunos de ellos puede ser muy complejo para un usuario que no tenga conocimientos mercantiles avanzados y hace que sea difícil de aprovechar si la pyme no tiene herramientas para capturar dicha información (crawlers, OCR, etc.), construir los sistemas de información, orquestar de manera eficiente el flujo de datos, optimizar rendimientos, o el personal adecuado para gestionar esa infraestructura. También sería necesario contar con científicos de datos que hagan la labor de exploración de la información, que sean capaces de darle sentido y además tengan una buena formación en temas de propiedad intelectual y de protección de datos (GDPR para sus siglas en ingles), porque no todos los datos que se pueden capturar se pueden utilizar. Hay pymes con un perfil muy tecnológico que se sienten cómodas con este tipo de tecnologías y tienen científicos de datos cualificados para trabajarlos, pero son una excepción. En cualquier caso, el coste de procesar esta información es muy elevado, por lo que salvo que la empresa use la información para construir productos o servicios basados en datos, tampoco tendría sentido realizar el esfuerzo de capturar y procesar dicha información, y de hacer el mantenimiento necesario para adaptarse a los continuos cambios en el formato y en los accesos a la información. Todo lo anterior hacen que ésta sea una opción inviable para la práctica totalidad de las empresas, ya sean pymes o grandes empresas.

- La información procedente de proveedores de datos externos que capturan, enriquecen, contrastan, limpian, normalizan y suministran los datos, para hacer que sean fácilmente explotables, suele ser la opción que mejor se ajusta a las necesidades de una empresa. Permite reducir el esfuerzo de captura e ingesta de la información, de forma que pueda centrarse en la explotación de la información. En cualquier caso, si la información se suministra

FIGURA 2
FUENTES DE DATOS MÁS HABITUALES



Fuente: Elaboración propia.

en bruto, la explotación del dato sigue siendo compleja para la mayor parte de las empresas y prefieren gestionarla a través de herramientas que también les ayuden a interpretar la información.

En definitiva, sin el apoyo de las tecnologías adecuadas o de herramientas de gestión del riesgo es prácticamente imposible para las Pymes lograr tener una visión integral, objetiva y actualizada de su cartera de clientes, hacer un seguimiento adecuado de su evolución, ni de integrar en sus procesos de decisión el ingente volumen de información que se genera sobre los clientes o proveedores en las diferentes áreas que interaccionan con ellos.

Volviendo al sector financiero, la amplia lista de tecnologías que hacen que las máquinas puedan percibir, comprender, actuar y aprender para ampliar las capacidades humanas le han permitido mejorar y automatizar muchos de sus procesos clave de negocio. Son también de las empresas que más están explorando las fuentes de datos alternativos (*alternative data*), para tratar de arañar algunos puntos de mejora a la capacidad predictiva de sus modelos, sobre todo en aquellos segmentos sobre los que no hay información o ésta es muy pobre.

No obstante, la información que alimenta esta nueva generación de modelos proviene de fuentes no estructuradas (tráfico web, aplicaciones digitales, interacciones con chatbots o con los servicios de atención al cliente, etc.) y han transformado su toma de decisiones y la precisión de sus modelos.

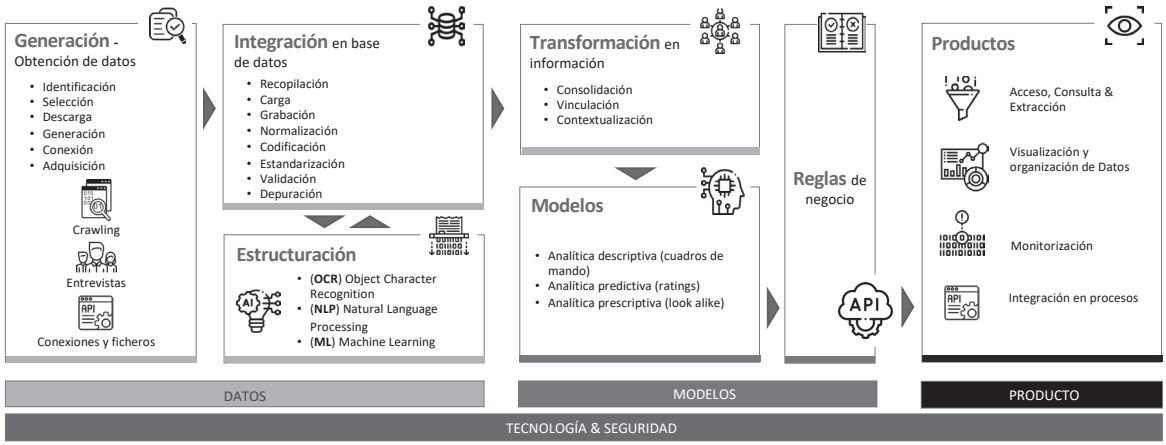
El resto del tejido empresarial debería gestionar su riesgo comercial con un enfoque similar al de una entidad financiera, evitando que los retrasos o impagos de sus clientes puedan afectar a su propia capacidad de pago o a la supervivencia de la propia empresa. No obstante, como hemos apuntado previamente, la realidad es bien distinta. Son pocas las Pymes que están aprovechando el potencial que aportan las nuevas tecnologías, cuando son precisamente el aliado que necesitan para desarrollar la capacidad de actuar de forma proactiva frente a los riesgos de impago y para implementar sistemas efectivos de seguimiento sobre la cartera de clientes como sí están haciendo las grandes empresas.

Esta clara falta de gestión y/o utilización de las herramientas adecuadas, provoca que miles de pequeñas y medianas empresas de toda Europa se vean abocadas al cierre, mientras esperan cobrar sus facturas. Estos cierres conllevan a su vez una importante destrucción de puestos de trabajo y suponen una tragedia personal para los empresarios implicados, algo que tampoco podemos olvidar a la hora de valorar el positivo impacto que tendría digitalizar y profesionalizar la gestión de riesgos en empresas de menor tamaño.

FIGURA 3
EJEMPLOS DE VISUALIZACIÓN AMPLIADO DEL RIESGO COMERCIAL DE UNA CARTERA DE CLIENTES

Fuente: Cuadro de mandos de la herramienta Insight View de Iberinform.

FIGURA 4
TRATAMIENTO DE LOS DATOS



Fuente: Elaboración propia.

CONSTRUCCION DE UNA SOLUCIÓN PARA LAS PYMES

A modo de ejemplo, a continuación, vamos a analizar cómo ha tratado de dar solución a esta problemática de las Pymes la herramienta Insight View (3) desarrollada por Iberinform y cómo se han aprovechado las tecnologías utilizadas por las grandes empresas para democratizar el acceso a soluciones avanzadas para la gestión de riesgos.

Esta solución captura diferentes fuentes de información estructurada y no estructurada, las procesa, limpia y normaliza y las utiliza para alimentar sus algoritmos predictivos y para ofrecer capas de visualización que facilitan su interpretación por parte de usuarios no avanzados. El usuario puede crear una lista de análisis de clientes o prospectos cargando un simple Excel y la herramienta en segundos genera un dash board que confirma al usuario el estado de cada empresa y su nivel de riesgo (ver imagen):

La herramienta ha encapsulado funcionalidades simplificadas de *Customer Analytics*, análisis de riesgos comerciales y un novedoso acceso a la información en modo DaaS (*Data as a Service*), permitiendo a más de 5.000 usuarios sin un conocimiento avanzado de análisis financiero y empresarial combinar la información interna de sus clientes con los datos de todas las empresas y autónomos de España y Portugal, o acceder online a datos de empresas en otros 72 países.

Son muchas las tecnologías empleadas en el *front-end* y en el *back-end* de Insight View, pero en este análisis nos vamos a centrar en las dos más relevantes: la migración a la nube (cloud) y la inteligencia artificial. La solución se lanzó en enero 2015 después de un exhaustivo análisis del mercado del segmento pymes y fue también la primera herramienta de este tipo que se apoyó en tecnologías cloud para, de una manera integrada a través de una única solución, ayudar a las empresas a buscar nuevos clientes, filtrar aquellos con menor riesgo comercial, conocer mejor la situación financiera de los principales y gestionar el riesgo de crédito de todos ellos de una manera ágil, sencilla e intuitiva.

Un elemento clave para el éxito del proyecto ha sido la escalabilidad y flexibilidad que aportaron las tecnologías cloud de Amazon Web Services (AWS), que permitieron dar servicio a miles de usuarios de forma concurrente, extraer mejor el valor de los datos y ofrecer un servicio más personalizado sin necesidad de complejas instalaciones ni de costosas inversiones, elementos que habían limitado previamente el despliegue de otras herramientas de la empresa. Esta infraestructura ha demostrado con el tiempo ser escalable, tolerante a fallos y tener una altísima disponibilidad. Es capaz de almacenar el histórico de datos de cerca de siete millones de empresas españolas, de conectarse con otros proveedores de datos empresariales para dar un acceso on line en 72 países, de permitir a los usuarios cargar sus propios datos y de realizar análisis sobre dichos datos de forma agregada con gran agilidad.

Además de visualizar de forma agregada los datos de una cartera de clientes o prospectos, la herramienta permite hacer un zoom sobre la situación de las empresas con mayor nivel de riesgo, consultar información de concursos de acreedores, incidencias judiciales, deudas con la Seguridad Social, la Agencia Tributaria, etc. así como consultar de forma ilimitada el fichero de morosidad bancaria Asnef Empresas y acceder al RAI.

TRATAMIENTO DE LOS DATOS

Los datos son obviamente la materia prima de la que se nutren este tipo de herramientas. Los podemos encontrar en diferentes fuentes y formatos y

requieren siempre de un tratamiento para que puedan aportar valor al negocio.

El primer paso en la cadena de valor del dato es la generación u obtención del mismo. En el caso de la herramienta Insight View, los datos se generan a través de entrevistas tanto telefónicas como presenciales con empresas y autónomos que operan en el mercado ibérico, a través de sistemas de *crawling* que descargan información de Internet, o a través de ficheros y conexiones directas con la Administración Pública o con otros proveedores de información.

Los datos que se capturan recogen diferentes aspectos de los agentes económicos como el número y tipología de empleados, oficinas y sucursales, situación financiera, incidencias de pago, evolución de las ventas, etc. obtenidos de fuentes públicas como el Registro Mercantil, Boletines Provinciales y Autonómicos, BOE, o el Registro Público Concursal entre otros, y de las propias empresas mediante entrevista.

Una vez obtenidos los datos, se les deben aplicar una serie de procesos para posibilitar su integración en una base de datos estructurada. Deben, por ejemplo superar determinadas validaciones que garanticen su calidad, o que identifiquen los errores que contienen para su depuración, han de acomodarse a unas estructuras establecidas que homogenicen el contenido, como por ejemplo ajustarse a un formato establecido de dirección postal, requieren habitualmente de una codificación definida que facilite su posterior explotación, etc. En los casos en los que los datos obtenidos son de tipo audio, imagen o texto, requieren además de un trabajo complejo de estructuración previo a su integración en la base de datos, como en el caso de los Informes de Auditoría que comentaremos más adelante.

Una vez que los datos dispersos, de diferentes fuentes y formatos, son consolidados en una base datos, podemos considerar que se han transformado en información, ya que un dato aislado probablemente apenas aporte valor, mientras que un dato localizable a través de un índice, vinculado a otros datos de la base, puesto en su contexto, y categorizado dentro de un conjunto con significado concreto, sí que es posible que aporte información a quien lo consulta.

Esta información hoy en día permite alimentar diferentes tipologías de modelos estadísticos que toman o ayudan a tomar decisiones de negocio. Desde la analítica descriptiva más básica, en la que mediante, por ejemplo, cuadros de mando se refleja la realidad de un sector económico; pasando por la analítica predictiva, en la que modelos estocásticos tratan de predecir el comportamiento de una determinada variable, como puede ser la solvencia futura de una empresa; hasta la analítica prescriptiva, en la que sistemas de optimización sugieren determinadas acciones, como la obtención de garantías en una operación comercial. Habitualmente el resultado en bruto que se obtiene de estos modelos se matiza con una serie de reglas de negocio que se establecen en base a la experiencia de expertos en la gestión de riesgos para mejorar su aplicación práctica.

La cadena valor básica del dato concluye con la explotación del mismo, mediante las herramientas o productos que permiten su acceso, consulta o extracción, su visualización integrada en cuadros de mandos o formando parte de determinados listados, o que facilitan su integración directa en otros procesos de negocio. Es importante destacar también la importancia del flujo proactivo de datos hacia los procesos de negocio, generando señales o enviando alertas sin esperar a la consulta por parte del usuario o del sistema implicado, lo que facilita la toma de decisiones adecuadas en el momento preciso.

Procesamiento automático de Informes de Auditoría

En el ámbito de la captura de información, los avances de la inteligencia artificial han supuesto una revolución, especialmente en el tratamiento de imágenes y textos.

Un ejemplo ilustrativo del gran avance que han supuesto estas nuevas tecnologías lo podemos encontrar en la captura y análisis de la información contenida en los Informes de Auditoría que las empresas depositan anualmente en el Registro Mercantil.

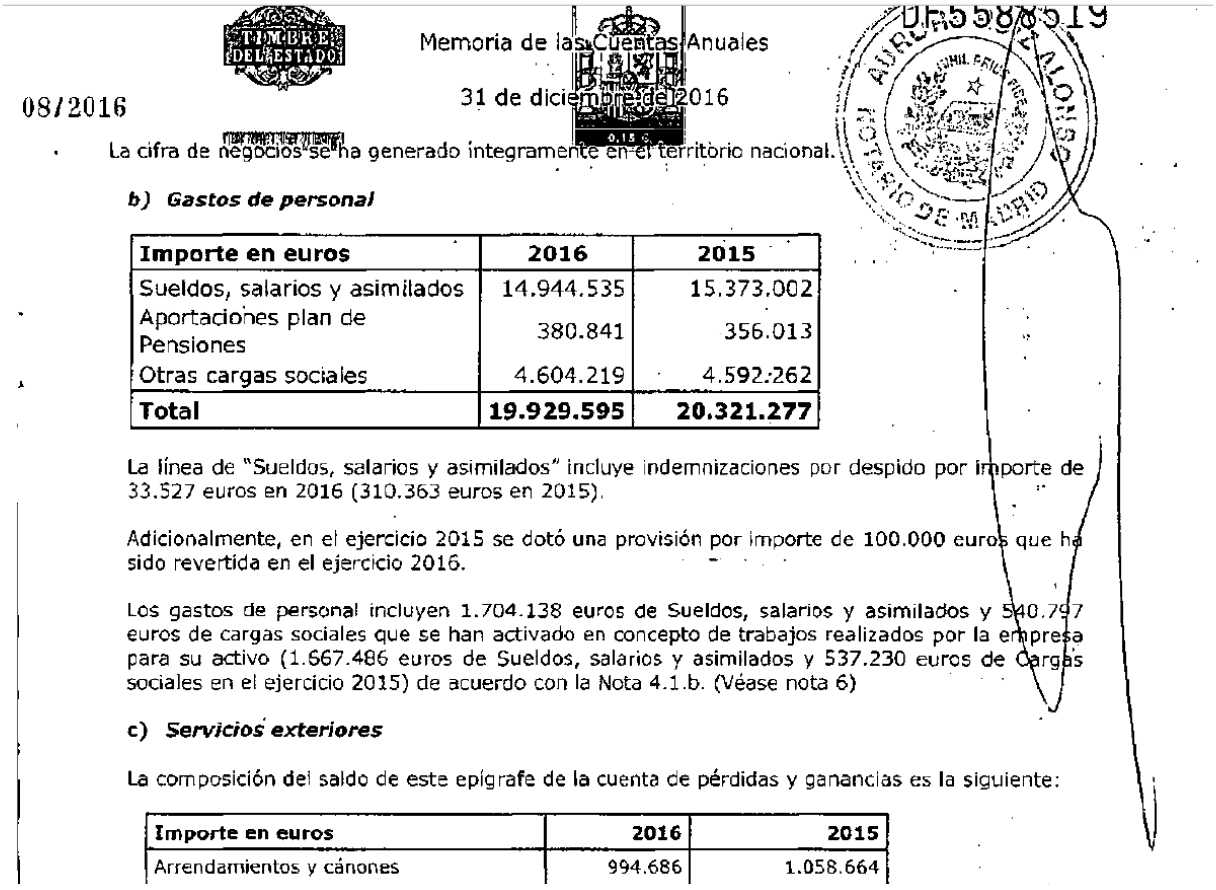
Naturaleza de la información

En las economías de mercado, las empresas están obligadas a depositar sus Cuentas Anuales en un Registro Público, consultable por cualquier agente que opere en el mismo. Esto es así porque en el intercambio económico la confianza es esencial y ésta se logra en gran medida con transparencia.

Las Cuentas Anuales son un conjunto de documentos que resumen la información financiera de una empresa y que engloban básicamente el balance, la cuenta de pérdidas y ganancias, el estado de cambios en el patrimonio neto y la memoria. En determinados casos, además se incorporan el estado de flujos de efectivo, y, sin formar parte de las mismas, se acompañan del Informe de Gestión, del Informe de Auditoría, del Informe de Autocartera (movimientos sobre acciones propias) y de la Página de Impacto Medioambiental.

Toda esta documentación es elaborada por la propia empresa, a excepción del Informe de Auditoría, que lo realiza un auditor externo tras llevar a cabo una revisión de los estados contables de la empresa para comprobar la exactitud de los apuntes contables y los criterios utilizados al realizarlos. Esta revisión tiene como fin asegurar que se han registrado todas las transacciones económicas y que los estados fi-

FIGURA 5
EXTRACTO DE MEMORIA DE LAS CUENTAS ANUALES



Memoria de las Cuentas Anuales

08/2016 31 de diciembre de 2016

La cifra de negocios se ha generado íntegramente en el territorio nacional.

b) Gastos de personal

Importe en euros	2016	2015
Sueldos, salarios y asimilados	14.944.535	15.373.002
Aportaciones plan de Pensiones	380.841	356.013
Otras cargas sociales	4.604.219	4.592.262
Total	**19.929.595**	**20.321.277**

La línea de "Sueldos, salarios y asimilados" incluye indemnizaciones por despido por importe de 33.527 euros en 2016 (310.363 euros en 2015).

Adicionalmente, en el ejercicio 2015 se dotó una provisión por importe de 100.000 euros que ha sido revertida en el ejercicio 2016.

Los gastos de personal incluyen 1.704.138 euros de Sueldos, salarios y asimilados y 540.797 euros de cargas sociales que se han activado en concepto de trabajos realizados por la empresa para su activo (1.667.486 euros de Sueldos, salarios y asimilados y 537.230 euros de Cargas sociales en el ejercicio 2015) de acuerdo con la Nota 4.1.b. (Véase nota 6)

c) Servicios exteriores

La composición del saldo de este epígrafe de la cuenta de pérdidas y ganancias es la siguiente:

Importe en euros	2016	2015
Arrendamientos y cánones	994.686	1.058.664

Fuente: Elaboración propia.

nancieros reflejan lo más fidedignamente posible la verdadera situación patrimonial y económica de la entidad y se presentan de acuerdo con los principios de contabilidad generalmente aceptados.

En su informe final de auditoría, el auditor expresa su opinión no vinculante sobre la fidelidad y exactitud de las cuentas examinadas. Se da por tanto el caso de que los datos presentados en determinadas Cuentas Anuales, a los ojos del auditor no reflejan fielmente la situación financiera de la empresa, y, por tanto, se niega a emitir una opinión, o la emite con salvedades que matizan las cuentas y derivado de ello, su interpretación y análisis.

Problemática de la explotación

Antes de la llegada de los sistemas de IA, Iberinform ya capturaba de forma automática determinados contenidos de las Cuentas Anuales, particularmente aquellos estructurados en origen, como el balance y la cuenta de pérdidas y ganancias. Esta información numérica se presenta tabulada y codificada, lo que hace relativamente sencilla su captura por parte un sistema OCR (*Optical Character Recognition*) de primera generación. Además, por lo general esta información se presenta ya en un formato estructurado (XBRL) (4), por lo que su integración en las bases de datos y su explotación en los procesos de negocio es directa y está ampliamente implantada.

No obstante, la información del Informe de Auditoría, al ser información no estructurada, debe ser capturada y analizada manualmente, lo que supone un alto coste de captura y por lo general genera un importante retardo en su incorporación a las aplicaciones de negocio. Esto provoca que determinados sistemas automáticos y modelos predictivos tomen decisiones basándose únicamente en los datos estructurados del balance o la cuenta de pérdidas y ganancias, sin tener en cuenta que en determinados casos un auditor ha planteado dudas sobre la veracidad o exactitud de los mismos.

La criticidad y urgencia de las decisiones derivadas del análisis de las Cuentas Anuales de una empresa, junto con el alto volumen de Informes de Auditoría que se publican anualmente, han justificado la aplicación de técnicas de Procesamiento del Lenguaje Natural (PLN) (5) para estructurar la información con-

tenida en los mismos e integrarla automáticamente en los procesos de negocio.

En línea con las conclusiones del artículo publicado en 2019 «¿Es útil el informe de auditoría externa para la predicción de quiebras? Evidencia utilizando inteligencia artificial» y el elaborado por profesores de la Universidad de Valencia Juan L. Gandía y David Huguet, «Textual y análisis de sentimientos en contabilidad», las conclusiones de TFM Delonix, presentado en mayo de 2018 en el Programa Ejecutivo en Big Data & Business Analytics de la Escuela de Organización Industrial (EOI), anticiparon la viabilidad técnica de procesar e interpretar esta información de forma masiva y la capacidad de la información contenida en los informes de auditoria para mejorar la capacidad predictiva de los modelos que anticipan la probabilidad de que una empresa incumpla con sus compromisos de pago, y sirvieron de base para acometer el desarrollo y puesta en producción de este proyecto.

El trabajo con imágenes

Por lo general, los modelos de PLN requieren que los datos de entrada al mismo sean en formato texto, por lo que, si la información no estructurada a tratar es un audio o una imagen, debe primero transformarse. Y es en este punto donde el tratamiento automatizado de la información de los Informes de Auditoría debe enfrentar un primer e importante escollo: la calidad de las imágenes.

Porque, aunque obviamente los informes son textos y hoy en día prácticamente cualquier texto se redacta utilizando un ordenador, los documentos de texto que constituyen o acompañan a las Cuentas Anuales se remiten al Registro Mercantil en formato imagen, pues deben imprimirse -habitualmente en papel timbrado, con marcas de agua o sellos- para ser firmados y posteriormente escaneados antes de presentarse.

Los resultados que los sistemas OCR de primera generación eran capaces de obtener sobre este tipo de imágenes eran relativamente pobres por la dificultad en aislar y reconocer los diferentes caracteres, al estar distorsionados, difuminados o poco definidos, superponerse a un sello, firma o marca, o confundirse con un fondo oscuro o con una perforación del papel, entre las múltiples fuentes de ruido.

Ha sido nuevamente la aplicación de la inteligencia artificial al reconocimiento de imágenes la que ha dado lugar a una nueva generación de OCRs que sí permiten hoy en día obtener unos resultados suficientemente satisfactorios como para facilitar a los modelos de procesamiento del lenguaje natural una materia prima con suficiente calidad.

No obstante, sigue siendo complejo identificar determinados errores en el procesamiento de las imágenes y siempre es necesario contar con un equipo de operadores de apoyo para la depuración de los mismos.

La extracción del Informe de Auditoría

Los diferentes documentos de texto que se remiten al Registro Mercantil (Memoria, Informe de Gestión, Informe de Auditoría, etc.) se presentan todos dentro de un único archivo de imagen, por lo que otra de las dificultades que es necesario superar es la de localizar y extractar aquellas páginas que componen el Informe de Auditoría.

Para ello los OCRs de última generación permiten identificar características especiales de los textos, como por ejemplo diferentes tamaños de letra o grosor para la identificación de epígrafes y otros metadatos sobre el texto que facilitan esta labor. No obstante, es necesario aplicar adicionalmente las primeras técnicas de PLN para localizar palabras clave que puedan indicar el inicio o finalización de la sección concreta a extractar.

Elementos a veces considerados secundarios, como los signos de puntuación, las mayúsculas y minúsculas o la estructuración en párrafos, también facilitan la aplicación de algoritmos de identificación de secciones y posteriormente son fundamentales en la captura de cifras numéricas o el reconocimiento de relaciones entre conceptos.

La estrategia de captura

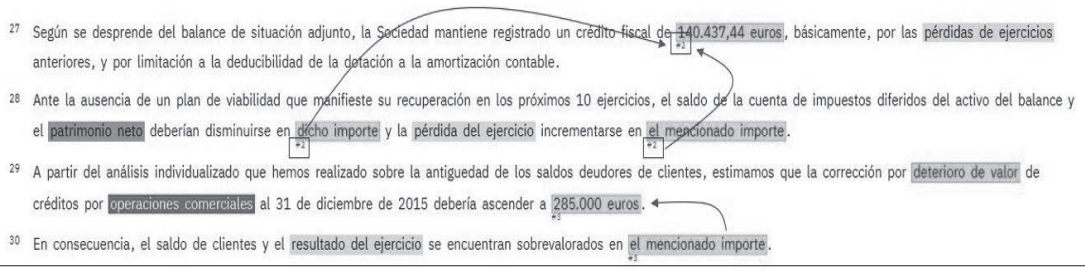
Una fase crucial de cualquier proyecto de procesamiento del lenguaje natural es la definición de la estrategia de captura, en la que es fundamental contar con la participación de expertos en la información a tratar.

A modo de ejemplo, en el caso concreto de los Informes de Auditoría, además de apoyarse en lingüistas computacionales, se contó con expertos en la elaboración de estos informes, es decir con auditores, y con analistas financieros, y especialistas en las decisiones de negocio que se toman basadas en la información contenida en los Informes de Auditoría.

Este equipo multidisciplinar de trabajo determinó que el objetivo de captura de información sería identificar quién era el auditor que firmaba el informe, qué opinión tenía de las cuentas, si percibía un riesgo en la continuidad del negocio, a qué partidas contables hacía referencia y qué matizaciones hacía sobre las mismas. Ello condujo a:

- Identificar los auditores acreditados para ejercer en España
- Categorizar las diferentes opiniones posibles
- Identificar palabras relacionadas con riesgo, duda, dificultad…

FIGURA 6
RELACIONES DE TEXTO



Fuente: Elaboración propia.

- Priorizar las partidas contables fundamentales
- Categorizar las posibles matizaciones sobre una partida contable

La obtención de entidades e importes

Una primera capa de procesamiento del lenguaje consiste en la localización de palabras o combinaciones de palabras concretas dentro del texto. Todas aquellas palabras o combinaciones de palabras que hacen referencia a un mismo concepto se agrupan en un diccionario o entidad. Por ejemplo, las palabras «inventario», «stock» o «existencias» se agrupan en el diccionario o entidad denominado «Existencias». Los diccionarios deben ser exhaustivos, recogiendo las diferentes grafías de las palabras, esto es, tanto en mayúsculas como en minúsculas, con y sin tilde, en singular y en plural, etc. para tratar de capturar todas las posibles formas de encontrar la palabra escrita en el texto.

También es necesario capturar y categorizar las cifras que aparecen en los textos, diferenciado importes, fechas, números de página, etc. La localización, por ejemplo, de importes que se ajustan a determinados patrones como dos cifras, guion, dos cifras, guion, y cuatro cifras, denotan fechas, por lo que la implementación de patrones de este tipo y sus variantes ayudan a tipificar las diferentes cifras de los textos.

Como se ha indicado anteriormente, la captura por el OCR de los signos de puntación en el caso de los importes es fundamental, puesto que por ejemplo no es lo mismo recoger la expresión «671 millones» que «67,1 millones». La combinación de cifras y textos, como en la expresión anterior o por ejemplo en fechas que combinan el número del día y el nombre del mes, también han de tenerse en consideración.

La implementación de estos patrones también se aplica a la combinación de elementos para construir expresiones regulares, siendo un ejemplo de este proyecto concreto la expresión «Importe moneda» que combinaba por un lado elementos de importes próximos a textos de moneda, tipo «671 millones de euros».

La construcción de los diferentes diccionarios o entidades, patrones básicos, y expresiones regulares, deriva de la estrategia de captura, y se realiza revisando cientos de ejemplos de Informes de Auditoría, con la intención de identificar y recoger las diferentes formas en las que se expresa un mismo un concepto.

Entrenamiento y relaciones

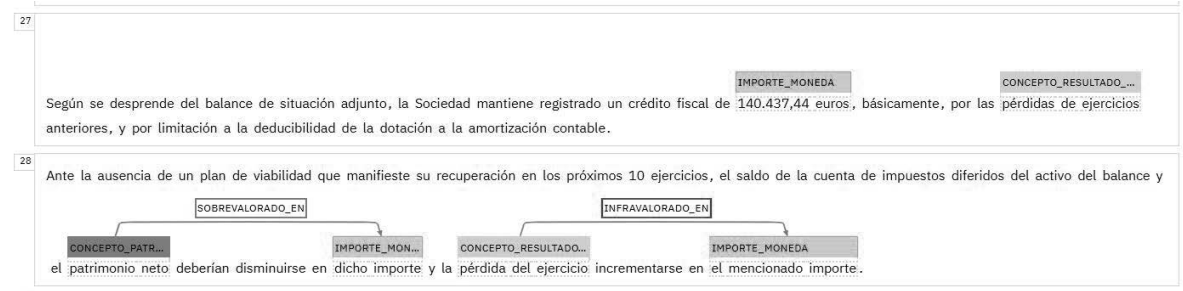
Las características de los textos que ayudan a delimitar las frases, como pueden ser los signos de puntuación o los párrafos facilitan también la captura de referencias cruzadas, como cuando en una misma frase se explicita un importe en cifras, y posteriormente se utiliza la expresión «dicho importe» o similar para referirse al mismo.

Además de las referencias cruzadas, es necesario recoger las diferentes relaciones entre entidades para lograr capturar el sentido de los textos. En la estrategia de captura se definieron los posibles tipos de relación entre las entidades que denotaban partidas contables y las expresiones de importe-moneda que las matizaban, como por ejemplo la relación «sobrevalorado en», o la relación «infravalorado en».

La construcción de las relaciones se realiza en base a técnicas de aprendizaje automático o Machine Learning (ML), en las que se entrena al modelo semántico a base de cientos de ejemplos que anotan manualmente los equipos del proyecto.

En la plataforma Watson Knowledge Studio de IBM, utilizada para el desarrollo de este proyecto, la anotación de los documentos se realiza de una manera sencilla e intuitiva, lo que mejora la agilidad y la velocidad con la que se preparan los ejemplos. El trabajo de anotación de un texto vendría a ser como marcar con un rotulador fluorescente de un color concreto todas las palabras que hagan referencia al mismo concepto. De esta forma se le indica a la máquina cómo poblar los diferentes diccionarios. Y la forma en la que se establecen las relaciones es arrastrando una entidad sobre aquella otra con la que esté relacionada y explicitando el tipo de relación entre las mismas.

FIGURA 7
PREANOTACIONES



Fuente: Elaboración propia.

Sobre todo en las primeras iteraciones de entrenamiento del modelo, es esencial que el equipo de anotación esté coordinado y discuta, consolide y acuerde los contenidos de los diccionarios y los criterios de captura de las diferentes relaciones para que todo el equipo anote de forma lo más uniformemente posible.

A medida que la máquina va aprendiendo a través de las diferentes iteraciones, la plataforma va facilitando el trabajo del equipo al ofrecer una «preanotación» en la que destaca a priori las identidades y relaciones que es capaz de localizar el modelo semántico, para que sólo sea necesario completarlas o corregirlas.

La nueva explotación de la información

Los principales beneficios de automatizar el procesamiento de los Informes de Auditoría son el ahorro de costes de grabación y la velocidad de incorporación de los datos del informe a los procesos de negocio.

Un sistema automatizado permite además abordar la totalidad de la fuente de información. Anteriormente, por el alto volumen de informes publicados, era habitual hacer una selección previa de las empresas más relevantes para el negocio y procesar únicamente los informes de las mismas.

Además, la captura manual por lo general se centraba en muy pocos campos, muy concretos y fáciles de localizar, dejando el trabajo de lectura amplia del texto a los analistas financieros. Esto generaba un importante cuello de botella en un recurso caro, como son los analistas, que debían secuenciar su trabajo retrasando la toma de decisiones de riesgo sobre empresas que, por otro lado, ya habían sido analizadas por otros sistemas automáticos (por ejemplo, calificadas por un modelo predictivo en base al balance y la cuenta de resultados) con información incompleta.

Si tenemos en cuenta que un altísimo porcentaje de los Informes de Auditoría apenas aportan información diferencial (simplemente corroboran la veracidad del contenido de las cuentas), es fácil concluir que el trabajo de los analistas era en gran medida similar a tener que «buscar una aguja en un pajar», lo que no favorece la priorización de los trabajos.

En concreto, el procesamiento automatizado de los Informes de Auditoría ha permitido:

- Procesar la totalidad de los Informes de Auditoría publicados en el Registro Mercantil
- Incorporar simultáneamente a los procesos de negocio la información de las Cuentas Anuales y la verificación de su validez por parte del auditor, facilitando la toma de decisiones con información completa.
- Corregir determinadas partidas contables en los valores indicados por el auditor para el recálculo de los modelos predictivos
- Incorporar a los modelos predictivos información no contenida en otras fuentes, y que se verificó que poseía un alto potencial predictivo.
- Generar alertas informativas sobre el contenido de los informes, facilitando la priorización de su análisis pormenorizado y evitando el trabajo sobre informes sin contenido.

Todo esto no habría sido factible abordarlo sin la aportación clave de los actuales sistemas basados en Inteligencia Artificial.

CONCLUSIÓN

La creciente cantidad de datos que las empresas generan hace que su almacenamiento, gestión y generación de conocimiento se hayan convertido en uno de los grandes retos para las organizaciones de todo el mundo.

Como hemos visto, la capacidad de las empresas para analizar el riesgo de impago de sus clientes es también un elemento crítico para su supervivencia y la mayoría de las pymes no tienen capacidad para abordar proyectos de IA con los que explotar la información y gestionar mejor sus riesgos, pero sí pueden

apoyarse en alguna de las herramientas de gestión de riesgos existentes en el mercado para dotarse de una mayor resiliencia ante las tensiones de tesorería que generan las crisis.

Soluciones de analítica para la gestión del riesgo como Insight View son un buen un ejemplo de cómo las nuevas tecnologías dan acceso a las Pymes a capacidades de análisis que antes eran accesibles sólo a grandes empresas, pero la mayor barrera para que estas interioricen la importancia que tiene una adecuada gestión del riesgo es sobre todo cultural. Una barrera en la que se entremezclan la falta de conocimientos básicos.

NOTAS

[1] Los diferentes tipos de empresa se clasifican por su tamaño y clasifica como pymes a las empresas con menos de 250 trabajadores y un volumen de negocio inferior a los 50 millones de euros, o bien aquellas que su balance general anual no sobrepasa los 43 millones de euros.

[2] Estudio que elaboran Iberinform y Crédito y Caución desde 2010 para analizar la gestión del riesgo comercial en España y los niveles de morosidad que experimentan las empresas.

[3] Herramienta de analítica desarrollada por Iberinform y lanzada al mercado en 2015.

[4] Conocido por su acrónimo XBRL (eXtensible Business Reporting Language), es el lenguaje, de libre uso, basado en los estándares XML, que permite la interoperabilidad y análisis de cualquier tipo de información financiera y empresarial a través de Intern*et al* integrar directamente las reglas de negocio en su desarrollo.

[5] En inglés, NLP, Natural Language Processing

REFERENCIAS

Altman, Edward I., and Thomas P. McGough (1974). Evaluation of a company as a going concern. Journal of Accountancy.

Abdou, H. A., & Pointon, J. (2011). Credit scoring, statistical techniques and evaluation criteria: A review of the literature. Intelligent Systems in Accounting, Finance and Management.

Guillermo Cabot y Lucas Suárez (2015). La gestión del Riesgo de crédito Comercial. LID Editorial.

Cenciarelli, Velia Gabriella, Giulio Greco, and Marco Allegrini (2018). External audit and bankruptcy prediction. Journal of Management and Governance.

Nora Muñoz Izquierdo, María del Mar Camacho Miñano, María Jesús Segovia Vargas y David Pascual Ezama (2019). Is the External Audit Report Useful for Bankruptcy Prediction? Evidence Using Artificial Intelligence. Complutense University of Madrid and Cunef. Spanish Accounting Review.

Textual analysis and sentiment analysis in accounting Juan L. Gandía, David Huguet, Departamento de Contabilidad, Facultad de Economía, Universidad de Valencia, Valencia, España.

Estudio de Gestión del Riesgo de Crédito en España (2021). Iberinform y Crédito y Caución.