

HERRAMIENTA DE PREDICCIÓN DE EMPRESAS AUDITABLES



< Conectando evidencias >

AUTORES:

ANTONIO NEGRÓN ROLDÁN
JOSÉ ANDRÉS LÓPEZ DE FEZ
JOSÉ MANUEL RODRÍGUEZ JIMÉNEZ
MANUEL CANO PAYÁN
MARTA ESQUIVEL RODRÍGUEZ
RAFAEL LUIS NAVIDAD MORALES

Proyecto: Herramienta de predicción de auditorías para empresas

Autores: ANR – JALF – JMR – MCP – MER – RNM

Versión: 2.0

INDICE

LISTA DE FIGURAS	iii
LISTA DE TABLAS	iv
HISTORIAL DE REVISIONES	v
1. Presentación.....	1
1.1 El Proyecto.....	1
1.2 El Equipo	2
1.3 ¿Por qué MLA ² ?	4
1.4 Misión, visión, lema.....	6
2. Definición del problema.....	7
2.1 La competencia	7
2.2 El mercado.....	8
2.3 El perfil de cliente.....	11
2.4 DAFO - CAME.....	14
2.5 Marco normativo.....	18
3. Modelo de negocio	21
3.1 Canvas.....	21
3.2 Caso de uso.....	23
3.3 Mapa estratégico.....	26
3.4 Beneficios esperados.....	29
4. Planes operativos	30
4.1 Plan de implantación y distribución	30
4.2 Plan de expansión.....	32
4.3 Plan de marketing	33
5. Solución tecnológica	34
5.1 Análisis funcional.....	35
5.2 Fuentes de datos	37
5.3 Ingesta de datos	39
5.4 Enriquecimiento y unificación.....	47
5.5 Modelo de predicción	48
5.6 Visualización de resultados	50
5.7 Arquitectura de la solución	52
5.8 Prototipo y validación	54
6. Estudio Económico-Financiero.....	63
6.1 Principales clientes y previsión de ingresos.....	63
6.2 Plan de inversiones, gastos iniciales y balance final previsional	65
6.3 Plan de financiación	67

	Programa Superior en Big Data & Business Analytics 7ªEd., Sevilla, Marzo – Septiembre 2020	
Proyecto: Herramienta de predicción de auditorías para empresas		
Autores: ANR – JALF – JMR – MCP – MER – RNM		Versión: 2.0

6.4	Cuentas de pérdidas y ganancias	69
6.5	Plan de Tesorería	71
6.6	Ratios y principales indicadores financieros	73
7.	Conclusiones	75

	Programa Superior en Big Data & Business Analytics 7ªEd., Sevilla, Marzo – Septiembre 2020	
Proyecto: Herramienta de predicción de auditorías para empresas		
Autores: ANR – JALF – JMR – MCP – MER – RNM		Versión: 2.0

LISTA DE FIGURAS

Figura 1: Equipo de proyecto.....	2
Figura 2: Logotipo de MLA ²	4
Figura 3: Resultado del ejercicio de identificación del cliente objetivo.....	12
Figura 4: Mapa de empatía del cliente objetivo.....	13
Figura 5: Canvas del modelo de negocio.....	22
Figura 6: Diagrama funcional del flujo de procesado de datos.....	35
Figura 7: Secuencia temporal del uso de los datos en el modelo predictivo y en la visualización.....	41
Figura 8: Modelo de datos de la tabla con todos los datos integrados.....	47
Figura 9: Pantalla inicial Módulo de visualización.....	50
Figura 10: Pantalla detalle empresa del Módulo de visualización.....	51
Figura 11: Arquitectura final.....	52
Figura 12: Pantalla resumen de una empresa, con indicación del número de empedados.....	54
Figura 13: Balance de los 3 últimos años de la empresa, desglosado por partidas principales.....	55
Figura 14: Cuenta de explotación de los 3 últimos años de la empresa, desglosado por partidas principales.....	56
Figura 15: Matriz de confusión del modelo de regresión logística.....	57
Figura 16: Métricas y curva ROC UAC del modelo de regresión logística.....	57
Figura 17: Matriz de confusión del modelo random-forest.....	58
Figura 18: Métricas del modelo random-forest.....	58
Figura 19: Curva ROC del modelo random forest con 1000 árboles.....	59
Figura 20: Peso de las 10 variables que justifican el 80% de la predicción.....	60
Figura 21: Matriz de confusión y métricas del modelo de ensembles con boosting....	61
Figura 22: Curva ROC AUC y métricas del modelo de ensembles con boosting.....	61
Figura 23: Balance previsional del MLA ² a 3 años.....	66
Figura 24: Cuenta de previsional del pérdidas y ganancias de MLA ² a 3 años.....	70
Figura 25: Saldo de tesorería y (necesidad financiera).....	71
Figura 26: Previsión de saldo de caja.....	71
Figura 27: Plan de tesorería del primer año.....	72

	Programa Superior en Big Data & Business Analytics 7ªEd., Sevilla, Marzo – Septiembre 2020	
Proyecto: Herramienta de predicción de auditorías para empresas		
Autores: ANR – JALF – JMR – MCP – MER – RNM		Versión: 2.0

LISTA DE TABLAS

Tabla 1: Paquetes de comercialización del servicio	24
Tabla 2: Tabla CdC.....	39
Tabla 3: Diccionario de la tabla de información microeconómica de las empresas Campos generales	42
Tabla 4: Diccionario de la tabla de información microeconómica de las empresas Campos del Balance anual	43
Tabla 5: Diccionario de la tabla de información microeconómica de las empresas Campos de la Cuenta de explotación anual	44
Tabla 6: Diccionario de las tablas de datos macroeconómicos	45
Tabla 7: Coordenadas de cada código postal español.....	45
Tabla 8: Campos adicionales a la macro-tabla del modelo predictivo	49
Tabla 9: Previsiones a 3 años de los principales ratios económicos y financieros de MLA ²	73

 Escuela de organización industrial	Programa Superior en Big Data & Business Analytics 7ªEd., Sevilla, Marzo – Septiembre 2020	
Proyecto: Herramienta de predicción de auditorías para empresas		
Autores: ANR – JALF – JMR – MCP – MER – RNM		Versión: 2.0

HISTORIAL DE REVISIONES

Versión	Fecha	Motivo de la versión
0.1	30.06.2020	Versión inicial: estructura básica
0.2	14.07.2020	Desarrollo de los capítulos 1, 2 y 5
0.3	21.07.2020	Desarrollo del capítulo 3
0.4	04.08.2020	Desarrollo del capítulo 4
0.5	18.08.2020	Revisión general capítulos 1, 2, 3 y 4
0.6	21.08.2020	Reescritura del capítulo 5 con nuevo paradigma de ingesta de datos
0.7	24.08.2020	Completada descripción de los apartados 5.5 y 5.6 Desarrollo del capítulo 6
1.0	25.08.2020	Revisión general de los capítulos 1, 2, 3, 4, 5 y 6. Primera versión completa
1.1	07.09.2020	Cambios diversos derivados del ensayo de la presentación con la clase
1.2	16.09.2020	Cambios diversos derivados del ensayo con el tutor
1.3	20.09.2020	Incorporación del apartado 6.5 Plan de tesorería
2.0	21.09.2020	Versión final

	Programa Superior en Big Data & Business Analytics 7ªEd., Sevilla, Marzo – Septiembre 2020	
Proyecto: Herramienta de predicción de auditorías para empresas		
Autores: ANR – JALF – JMR – MCP – MER – RNM		Versión: 2.0

1. Presentación

1.1 El Proyecto

El objetivo de MLA² es incrementar la cartera comercial de nuestros clientes, las pequeñas firmas locales de auditoría.

Con el apoyo de una plataforma basada en *Machine Learning* y *Business Analytics*, se analizarán datos fundamentalmente estructurados de las empresas ubicada en el área de influencia de cada uno de nuestros clientes.

De esta manera, nuestros clientes podrán hacer uso de la información privilegiada que nuestro sistema de predicción les proporcionará para anticiparse al momento en que una empresa alcanza a cumplir los requisitos legales para ser auditada.

Esto otorgará a nuestros clientes la capacidad de gestionar la iniciativa en su acercamiento comercial a sus clientes potenciales y así gozar de una ventaja competitiva.

1.2 El Equipo

El equipo de MLA² lo compone un grupo de profesionales provenientes de diferentes sectores y con muy diverso *background* tanto académico como laboral, pero a los cuales les une la pasión por el análisis de los datos y sobre todo un objetivo común, el hacer que la tecnología reduzca la brecha competitiva que existe entre las grandes y las pequeñas empresas.

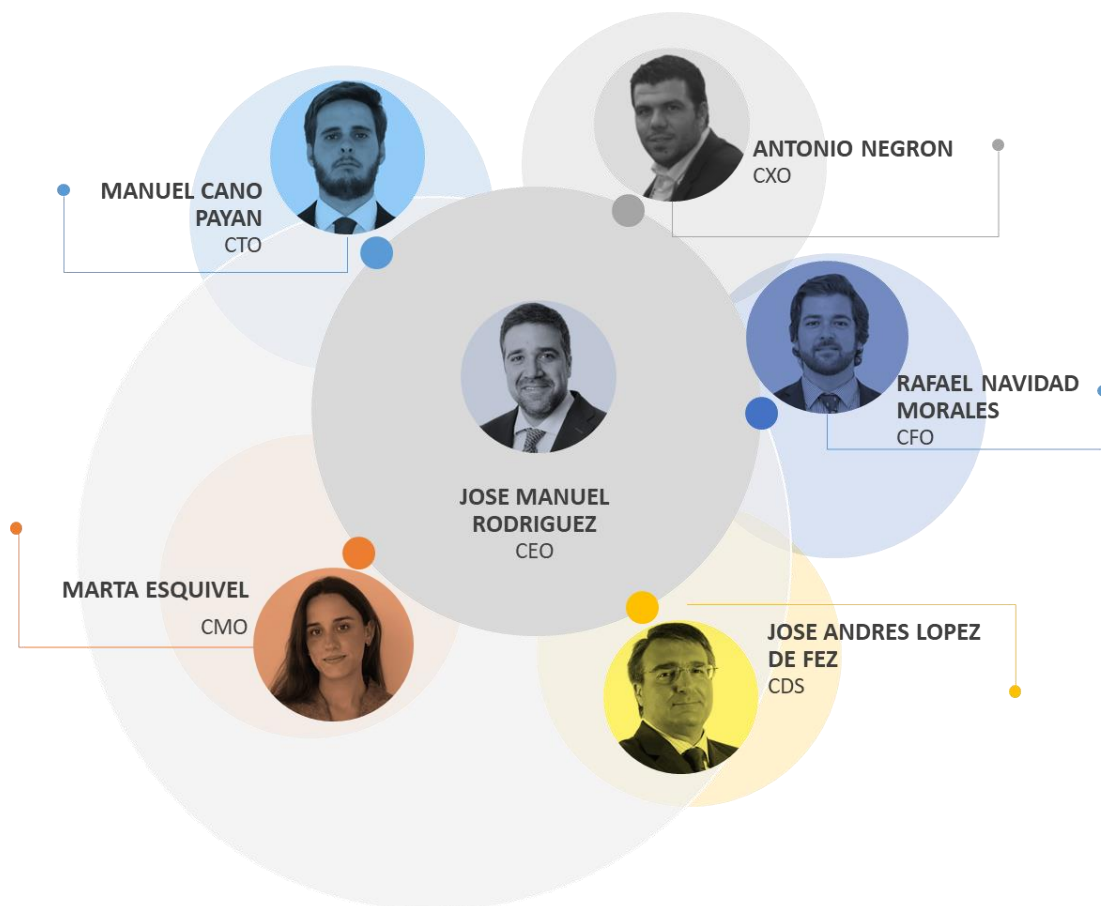


Figura 1: Equipo de proyecto

	Programa Superior en Big Data & Business Analytics 7ªEd., Sevilla, Marzo – Septiembre 2020	
Proyecto: Herramienta de predicción de auditorías para empresas		
Autores: ANR – JALF – JMR – MCP – MER – RNM		Versión: 2.0

CEO (*Chief Executive Officer*) – José Manuel Rodríguez Jiménez (JMR)

Background: CIO de multinacionales del sector de la ingeniería

CFO (*Chief Financial Officer*) – Rafael Navidad Morales (RNM)

Background: Auditor Senior en grandes firmas a nivel nacional

CDS (*Chief Data Scientist*) – José Andrés López de Fez (JALF)

Background: Director de Operaciones y Calidad en multinacional vinculada a la defensa.

CMO (*Chief Marketing Officer*)– Marta Esquivel Rodríguez (MER)

Background: Auditor en una de las Big Four

CTO (*Chief Technology Officer*) – Manuel Cano Payán (MCP)

Background: Business Analyst en grandes consultoras internacionales

CXO (*Chief Experience Officer*) – Antonio Negrón Roldan (ANR)

Background: Consultor experto en planificación en grandes firmas nacionales.

	Programa Superior en Big Data & Business Analytics 7ªEd., Sevilla, Marzo – Septiembre 2020	
Proyecto: Herramienta de predicción de auditorías para empresas		
Autores: ANR – JALF – JMR – MCP – MER – RNM		Versión: 2.0

1.3 ¿Por qué MLA²?

MLA² surge con la idea de aunar términos relacionados con *Big Data* y auditoría. El proceso de búsqueda del nombre comenzó buscando juegos de palabras que tuvieran un sentido con nuestra idea de negocio, que no es otro detectar la necesidad de auditoría de una empresa apoyándonos para ello en una plataforma basada en *Machine Learning*.

Con ello surge MLA², *Machine Learning Audit and Accounting*, cuyas siglas serían MLAA y cuya doble A se acordó simbolizar con una A elevada al cuadrado, operación que evoca a las matemáticas y la ciencia que hay detrás de los servicios ofertados.

Es un nombre corto y fácil de recordar para así conseguir atraer a nuevos clientes.



Figura 2: Logotipo de MLA²

En la imagen corporativa se ha optado por la sencillez y neutralidad. En el mundo de la auditoría y también en la época en la que vivimos el logotipo de las grandes empresas se ha ido transformando, fomentándose una imagen moderna y sofisticada, evitando las relaciones con objetos físicos y reales, así como con nombres propios. Creando una marca para todos que se mantenga en el tiempo.

Partiendo de lo anterior, se ha buscado en todo momento encajar el logo en el sector auditor, con el fin de que nos ayude a posicionar MLA² ya que la imagen puede ser un freno cuando la es demasiado recargada o pasada de moda.

Se han utilizado formas simples y minimalistas, las líneas rectas indican solidez. En lo referente a los colores, se han seleccionado el gris y el naranja. El naranja es un color

	Programa Superior en Big Data & Business Analytics 7ªEd., Sevilla, Marzo – Septiembre 2020	
Proyecto: Herramienta de predicción de auditorías para empresas		
Autores: ANR – JALF – JMR – MCP – MER – RNM		Versión: 2.0

con fuerza, llamativo con gran visibilidad que capta la atención y destaca aspectos, mientras que el gris implica seguridad y fiabilidad, siendo un color de fondo que contrasta con el naranja. Funciona y se mezcla bien con cualquier otro color, siendo muy utilizado en el sector tecnológico.

	Programa Superior en Big Data & Business Analytics 7ªEd., Sevilla, Marzo – Septiembre 2020	
Proyecto: Herramienta de predicción de auditorías para empresas		
Autores: ANR – JALF – JMR – MCP – MER – RNM		Versión: 2.0

1.4 Misión, visión, lema

Nuestra misión:

Ayudar a las pequeñas firmas de auditoría españolas a acercarse con anticipación a sus potenciales clientes, optimizando su planificación y gestión de recursos.

Nuestra visión:

Consolidarnos como el *partner* tecnológico para la transformación digital de las empresas de auditoría europeas, y convertirnos en el referente de sistemas financieros predictivos.

Nuestro lema:

“Conectando evidencias”

Evidencias es la palabra clave que un auditor usa en su trabajo diario. Con este juego de palabras definimos MLA², como la empresa que, usando evidencias, las obtenidas por nuestro sistema predictivo, es capaz de proporcionar a sus clientes una ventaja competitiva clave.

	Programa Superior en Big Data & Business Analytics 7ªEd., Sevilla, Marzo – Septiembre 2020	
Proyecto: Herramienta de predicción de auditorías para empresas		
Autores: ANR – JALF – JMR – MCP – MER – RNM		Versión: 2.0

2. Definición del problema

2.1 La competencia

Habiendo realizado un análisis de la potencial competencia existente en el mercado se concluye que no existe en la actualidad ninguna empresa que preste el servicio ofrecido por MLA².

Hay empresas, como elnforma o Infocif, con las que MLA² cuenta, que serán los proveedores de datos para la prestación del modelo de servicio ofrecido por MLA², y que potencialmente podrían dar el salto a prestar un servicio similar pero actualmente no hay evidencias de ningún interés por parte de estas compañías de dar el salto a ese tipo de servicio.

De cualquier modo, en el apartado 2.4 se aborda el DAFO del proyecto y se analizan los posibles planes de acción en caso de que aumente la probabilidad de que este riesgo se materialice.

	Programa Superior en Big Data & Business Analytics 7ªEd., Sevilla, Marzo – Septiembre 2020	
Proyecto: Herramienta de predicción de auditorías para empresas		
Autores: ANR – JALF – JMR – MCP – MER – RNM		Versión: 2.0

2.2 El mercado

Los clientes de MLA² son pequeñas y medianas auditoras que tienen el objetivo de expandirse incorporando tecnología al negocio.

El objetivo es dotar de información personalizada a estas auditoras para que puedan anticiparse a la competencia aumentando su cartera de clientes.

En España hay más de 20.000 auditores y 1.300 firmas de auditoría.

Se emiten al año más de 62.000 informes, fundamentales para, entre otros aspectos, la seguridad de los inversores en miles de empresas, su buen gobierno corporativo y la protección de los derechos de los accionistas.

Es, por tanto, que la auditoría juega un papel muy relevante tanto en el desarrollo económico como en la transparencia empresarial.

Como indica el siguiente texto extraído de “El Portal del auditor¹”, el sector de la auditoría es fundamental para la credibilidad de la economía en general.

“Por estas razones, el Instituto de Censores Jurados de Cuentas de España (ICJCE) y el Consejo General de Economistas desean poner de manifiesto la relevancia que la auditoría tiene para el desarrollo económico, especialmente en un momento en el que la Sociedad española demanda más transparencia a las instituciones públicas y privadas. Que sus trabajos sean rigurosos, y sean comprendidos y respetados no sólo es necesario para los auditores, también lo es para el conjunto de los ciudadanos. Afortunadamente, la mayoría de la sociedad es consciente de que el trabajo del auditor resulta fundamental para aumentar la fiabilidad y credibilidad de la información financiera, como revela que en 2011 –en el punto álgido de la crisis- los trabajos de la auditoría de carácter voluntario representaran un 27% del total de las auditorías realizadas”.

Según Marco Trombetta del Instituto Empresa, en su artículo “*Small is big! The role of ‘small’ audits for studying the audit market*” el 95% de las firmas auditoras en España son PYMES.

La competencia entre las auditoras es muy intensa y, aunque las grandes compañías se encuentren auditadas por las conocidas *Big Four*, existen muchas medianas

¹ <https://www.icjce.es/home/>

	Programa Superior en Big Data & Business Analytics 7ªEd., Sevilla, Marzo – Septiembre 2020	
Proyecto: Herramienta de predicción de auditorías para empresas		
Autores: ANR – JALF – JMR – MCP – MER – RNM		Versión: 2.0

empresas que por sus características deben ser auditadas necesitando de un auditor de confianza.

Las denominadas *Big Four* (Deloitte, PWC, EY y KPMG) especializan sus servicios en las empresas cotizadas, y por su estructura y dimensión no son competitivas a la hora de auditar el resto de las compañías del tejido empresarial español, donde debido a la existencia de multitud de estas firmas de auditoría PYMES los precios tienden a un equilibrio de mercado.

La conclusión es que existe una competencia muy alta en el sector de las pequeñas firmas auditoras, donde el precio es clave y por tanto los márgenes muy ajustados y donde es fundamental ser el primero en llegar al cliente.

Como se ha mencionado anteriormente, la idea de negocio de MLA² se centra en medianas auditoras. Estos serán nuestros clientes principales. Más concretamente, serán auditoras que buscan aumentar su cartera de clientes o que debido a la normativa las empresas a las que auditan se ven en la necesidad de cambiar de auditor.

En la práctica, son las auditoras las que dedican mucho tiempo a la búsqueda de potenciales clientes y en las elaboraciones de propuestas de auditoría.

IFAC² (*International Federation of Accountants*) realiza anualmente una encuesta a medianas y pequeñas auditoras sobre sus desafíos y visión del futuro en el sector. Ya en la encuesta realizada en 2014 aparecía como una de las principales preocupaciones la atracción de nuevos clientes para los auditores de medianas y pequeñas empresas (PYMES), más concretamente un 59% de los encuestados en Europa lo calificaban como un desafío alto. En 2018 ya se comprueba que estos auditores están adoptando la tecnología para mejorar su servicio y a atraer y retener.

La última encuesta a la que se ha podido tener acceso es la de 2018³, en la que participaron 150 países. En dicha encuesta se observa que el 28% de auditoras planea asignar más del 10% de sus ingresos totales a inversión en tecnología, esto es apreciado como un desafío significativo. Además, el 46% ve como una de sus principales oportunidades de innovación la atracción y retención de nuevos clientes. Es por ello que su solución estaría en MLA².

² <https://www.ifac.org/>

³ Fuente: <https://incp.org.co/Site/publicaciones/info/archivos/Global-SMP-2018-Summary-30102018.pdf>

	Programa Superior en Big Data & Business Analytics 7ªEd., Sevilla, Marzo – Septiembre 2020	
Proyecto: Herramienta de predicción de auditorías para empresas		
Autores: ANR – JALF – JMR – MCP – MER – RNM		Versión: 2.0

Un análisis del Registro Oficial de Auditores de Cuentas⁴ ha permitido detectar que más de un 50% de las firmas están presentes únicamente en una provincia por lo que el nivel de ofertas que pueden recibir las empresas es muy amplio, teniendo en cuenta que muchas auditoras no se limitan a las provincias si no que aumentan su área de trabajo más allá de las provincias limítrofes.

El foco de estudio y de análisis de MLA² se centra en las empresas no muy grandes, ya que éstas seguramente buscarán una gran auditora que pueda abarcar con sus infraestructuras el análisis necesario en una auditoría de cuentas. Actualmente en la composición del sector empresarial las PYMES abarcan un 95%.

⁴ <http://www.icac.meh.es/ROAC/ROAC.aspx>

	Programa Superior en Big Data & Business Analytics 7ªEd., Sevilla, Marzo – Septiembre 2020	
Proyecto: Herramienta de predicción de auditorías para empresas		
Autores: ANR – JALF – JMR – MCP – MER – RNM		Versión: 2.0

2.3 El perfil de cliente

En base a todo lo anterior, hay que tener en cuenta que existe una competencia muy alta en el sector de las medianas firmas auditoras, donde el precio y la anticipación son claves, los márgenes son muy ajustados y prima ser el primero en llegar al cliente.

Como hemos comentado en puntos anteriores, el objetivo es incrementar la cartera comercial de nuestros clientes, siendo estos las medianas y pequeñas auditoras locales, basándonos en el apoyo de una plataforma basada en *Big Data* y *Machine Learning* para analizar datos estructurados y no estructurados de los mercados locales, para que así nuestro cliente pueda anticiparse a sus competidores al dotarles de información personalizada.

Para poder centrar bien el proyecto, se ha realizado un análisis y se ha configurado un esbozo de nuestro cliente potencial, dándole forma a ponerle nombre y definiendo sus características.

- **Cliente:** pequeñas y medianas auditoras que tienen el objetivo de expandirse incorporando tecnología al negocio.
- **Ámbito:** Nacional. En España hay más de 3.000 sociedades auditoras repartidas por toda España, aunque se identifican zonas con una competencia más intensa, como Madrid, Barcelona, Valencia y Sevilla.

	Programa Superior en Big Data & Business Analytics 7ªEd., Sevilla, Marzo – Septiembre 2020	
Proyecto: Herramienta de predicción de auditorías para empresas		
Autores: ANR – JALF – JMR – MCP – MER – RNM		Versión: 2.0

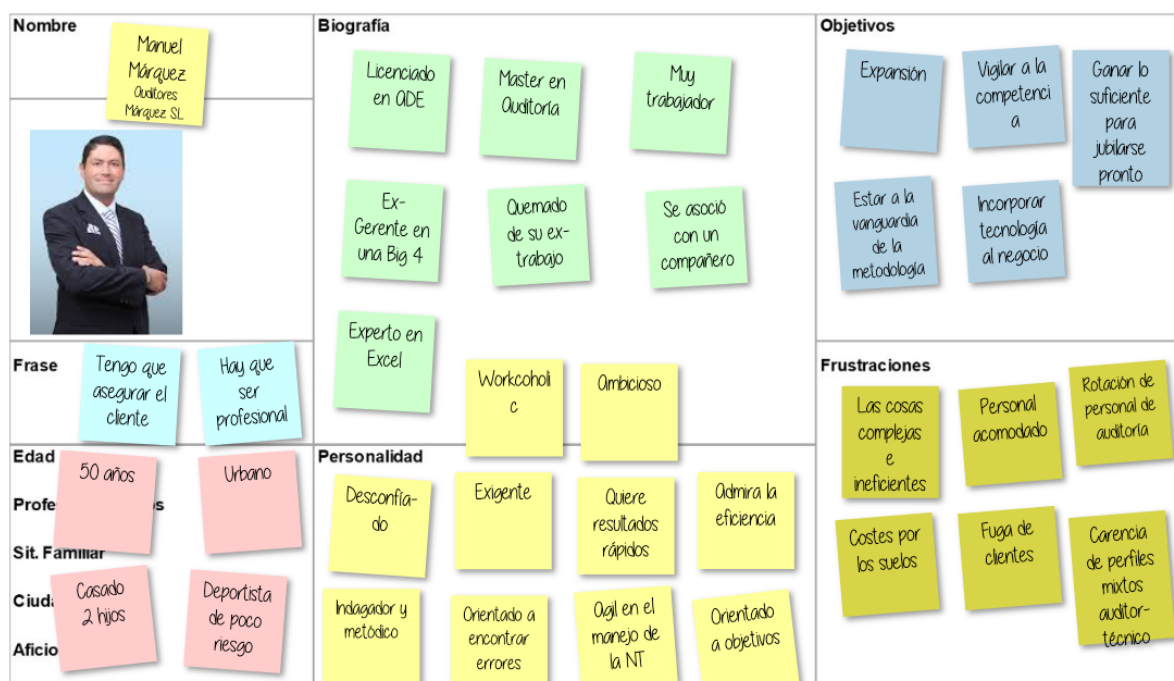


Figura 3: Resultado del ejercicio de identificación del cliente objetivo

En el cuadro anterior, se puede ver el resultado del proceso de identificación del perfil general que define a un auditor prototipo de nuestro cliente objetivo. El objetivo es conocerle y ver el problema que tiene, para poder definir de forma precisa la propuesta de valor que se le va a ofrecer.

La principal conclusión es que se trata de una persona exigente que quiere obtener resultados rápidos y eficiente. Además, es una persona desconfiada que suele detectar los errores.

Su objetivo como dueño de una auditora es la expansión incorporando nueva tecnología al negocio estando a la vanguardia de la tecnología. Le frustran las cosas complejas e ineficiente, así como no poder soportar los costes y la fuga de clientes entre otras preocupaciones.

La frase que le define no es otra que *“tengo que ser profesional y asegurar al cliente”*. Por lo que podemos comprobar que el cliente, la tecnología y los costes están dentro de sus preocupaciones y puntos importantes en la auditoría.

Indagamos más en las preocupaciones de nuestro cliente para lo cual hacemos uso del mapa de empatía.

Proyecto: Herramienta de predicción de auditorías para empresas

Autores: ANR – JALF – JMR – MCP – MER – RNM

Versión: 2.0

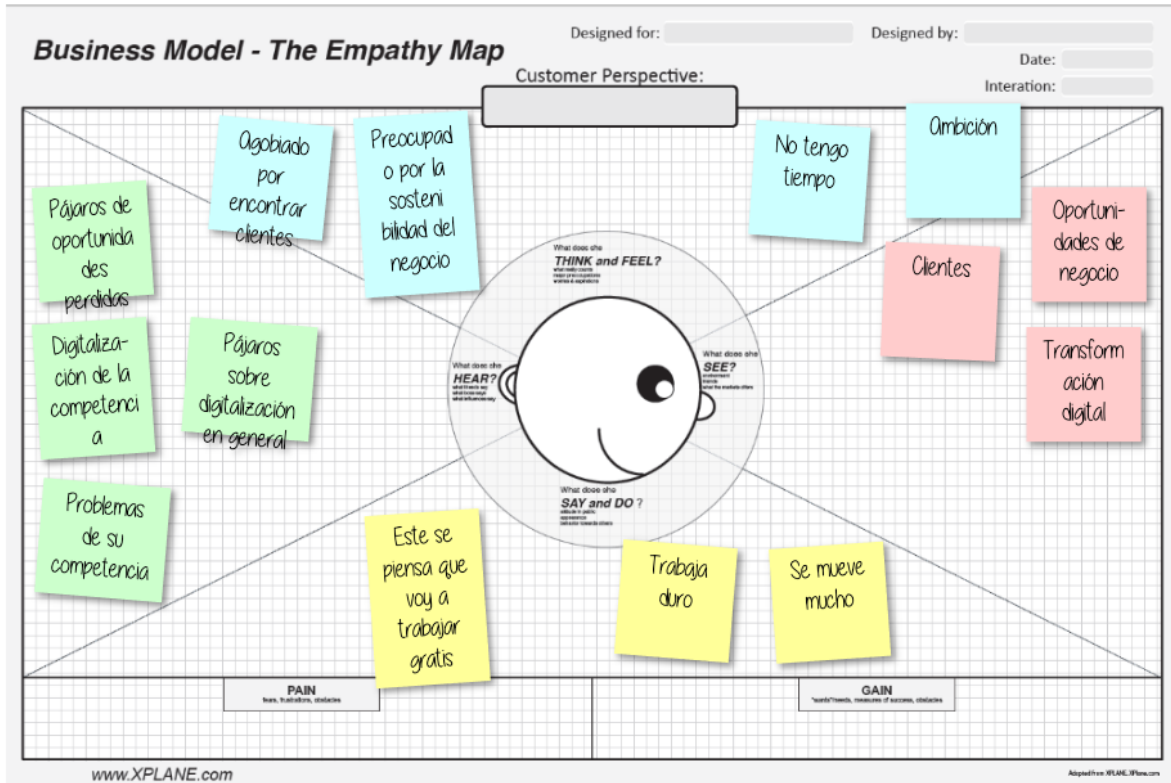


Figura 4: Mapa de empatía del cliente objetivo

En el mapa se identifican los cuatro aspectos principales de preocupación de nuestro auditor:

- **Qué dice/hace:** el sector se mueve mucho y trabaja duro.
- **Qué oye:** oportunidades perdidas, el avance de la competencia en la digitalización.
- **Qué ve:** clientes y oportunidades de negocio, además de transformación digital.
- **Qué siente:** se siente agobiado por encontrar clientes y por el tiempo, así como por la sostenibilidad del negocio.

 Escuela de organización industrial	Programa Superior en Big Data & Business Analytics 7ªEd., Sevilla, Marzo – Septiembre 2020	
Proyecto: Herramienta de predicción de auditorías para empresas		
Autores: ANR – JALF – JMR – MCP – MER – RNM		Versión: 2.0

2.4 DAFO - CAME

Se presenta a continuación el análisis DAFO – CAME realizado.

DAFO

- **Debilidades**

- El modelo económico es dependiente del coste de obtención de los datos y por tanto de fuentes de datos de terceros.
- No es sencillo incluir variables macroeconómicas que fortalezcan el modelo a la hora de predecir tendencias en las principales variables explicativas del modelo.
- Aunque el equipo tiene una sólida base teórica y tecnológica, no goza de una amplia experiencia en proyectos de *Big Data* y disciplinas asociadas como *Machine Learning* o *Business Analytics*.

- **Amenazas**

- Como se ha comentado anteriormente, las empresas de fuentes de datos pueden tener fácil la replicación de la idea explotando sus propios datos.
- El posible impacto de crisis sobrevenidas en periodos interanuales es difícilmente modelable.
- Existe en el sector una cierta tendencia expansiva de las *Big Four* hacia empresas auditadas de menor tamaño, lo que puede menoscabar el potencial de nuestro principal mercado.
- Esta concentración del sector de auditoría es una amenaza en la medida en que pueden tener una cuota de mercado que haga poco interesante nuestro servicio.

	Programa Superior en Big Data & Business Analytics 7ª Ed., Sevilla, Marzo – Septiembre 2020	
Proyecto: Herramienta de predicción de auditorías para empresas		
Autores: ANR – JALF – JMR – MCP – MER – RNM		Versión: 2.0

- **Fortalezas**

- Se ofrece un servicio altamente especializado.
- Dentro del equipo hay un alto conocimiento interno del sector de auditoría al proceder de este dos de los miembros.
- El modelo de negocio destaca por su simplicidad.
- Aunque la proporción de costes fijos frente a variables es alta, estos son bajos en comparación con los ingresos.

- **Oportunidades**

- Como se ha comentado anteriormente, no existe en la actualidad este servicio en el mercado, por lo que no hay competencia.
- En los contactos realizados en el sector se detecta interés por un servicio como el ofrecido por MLA².
- MLA² puede ayudar a la *democratización* del sector al facilitar el desarrollo de negocio de cualquier firma de auditoría, independientemente de su tamaño.
- La concentración del sector de auditoría que se está produciendo se detecta como una oportunidad para MLA² en la medida en que puede ayudar a potenciar las sinergias de la concentración.
- Hay un número elevado de clientes potenciales (casi 4.000).
- El modelo de negocio presenta una alta escalabilidad potencial como, por ejemplo:
 - Facilitando otras predicciones con similares fuentes de datos y clientes para casos como el de las licitaciones públicas, presentación de informes financieros o crediticios, etc.
 - Posible creación de un *marketplace* de auditores para servicios de auditoría y otros asociados.

 Escuela de organización industrial	Programa Superior en Big Data & Business Analytics 7ªEd., Sevilla, Marzo – Septiembre 2020	
Proyecto: Herramienta de predicción de auditorías para empresas		
Autores: ANR – JALF – JMR – MCP – MER – RNM		Versión: 2.0

CAME

- **Corregir:**

- Es posible negociar acuerdos por volumen de consultas con los proveedores de datos y al mismo tiempo es posible analizar a fondo el modelo predictivo para reducir, al mínimo indispensable, la necesidad de datos.
- Buscar y negociar con proveedores alternativos para aumentar la variedad de proveedores de datos.
- Monitorizar periódicamente el coste de las fuentes de datos.
- Vinculación con proveedores de primer nivel.
- Incluir estudios de mercado sectoriales y regionales.
- Entrenar el modelo en base a datos históricos de variables macroeconómicas como el PIB, IPC, Tasa de Empleo, etc.

- **Afrontar:**

- Vendernos como un servicio especializado en el sector de la auditoría: *“por auditores, para auditores”*.
- Realizar una buena campaña de *marketing / networking* para comunicar las bondades de ser auditado de cara a obtener mejor financiación. Es decir, realizar campañas de comunicación dirigidas no sólo a auditores sino a empresas auditables.
- Intentar mantenerse al día en cuanto a los movimientos en el mercado de empresas de auditoría (fusiones o adquisiciones).

- **Mantener:**

- Intentar mantenerse al día del mercado en cuanto a los movimientos en el mercado de empresas de auditoría (fusiones o adquisiciones).
- Hacer seguimiento constante de nuestros clientes.
- Mantener siempre al menos a un auditor en plantilla.

	Programa Superior en Big Data & Business Analytics 7ªEd., Sevilla, Marzo – Septiembre 2020	
Proyecto: Herramienta de predicción de auditorías para empresas		
Autores: ANR – JALF – JMR – MCP – MER – RNM		Versión: 2.0

- Foco en el negocio *core*.
- Mantener una estrategia de contención de los costes fijos, maximizando la relación coste variable vs fijo.
- Tener una monitorización de los costes de los proveedores de tecnología y de la evolución del mercado de datos.
- Monitorización de la tecnología en sí que soporta nuestro modelo de negocio.
- **Explotar**
 - Probar un “*producto mínimo viable*” (PMV) con rapidez antes de lanzar el servicio a escala nacional.
 - Realizar una importante acción de marketing y comunicación a nivel nacional una vez validado dicho producto mínimo viable.
 - Trabajar el posicionamiento de la marca desde el primer momento (servicio especializado, de auditor a auditor, fiabilidad, confianza, apoyo al pequeño auditor, etc.)
 - Organización de *workshops* con los clientes para identificar necesidades de servicios adicionales. Es decir, evolucionar nuestro servicio de la mano de nuestros clientes.

	Programa Superior en Big Data & Business Analytics 7ªEd., Sevilla, Marzo – Septiembre 2020	
Proyecto: Herramienta de predicción de auditorías para empresas		
Autores: ANR – JALF – JMR – MCP – MER – RNM		Versión: 2.0

2.5 Marco normativo

La disposición adicional primera de la Ley 22/2015, de 20 de julio, de Auditoría de Cuentas⁵ establece los supuestos que de cumplirse ocasionarán que las cuentas anuales deban ser auditadas de forma obligatoria. No obstante, a la hora de establecer magnitudes cuantitativas este precepto hace referencia a otras normas, como el Reglamento de Auditoría o la Ley de Sociedades de Capital, que pueden hacer difícil conocer de forma exacta si una empresa está obligada o no a auditar sus cuentas anuales.

En este sentido, aunando toda la normativa sobre el tema, se puede hacer una lista exhaustiva sobre las circunstancias que hacen que la auditoría de cuentas sea obligatoria:

- Cuando la entidad emita valores admitidos a negociación en mercados secundarios oficiales de valores o sistemas multilaterales de negociación.
- Cuando la entidad emita obligaciones en oferta pública.
- Si la entidad se dedica de forma habitual a la intermediación financiera, y, en todo caso, las entidades de crédito, las empresas de servicios de inversión, las sociedades rectoras de los mercados secundarios oficiales, las entidades rectoras de los sistemas multilaterales de negociación, la Sociedad de Sistemas, las entidades de contrapartida central, la Sociedad de Bolsas, las sociedades gestoras de los fondos de garantía de inversiones y las demás entidades financieras, incluidas las instituciones de inversión colectiva, fondos de titulización y sus gestoras, inscritas en los correspondientes Registros del Banco de España y de la Comisión Nacional del Mercado de Valores.
- Si la entidad tiene por objeto social cualquier actividad sujeta al Texto Refundido de la Ley de ordenación y supervisión de los seguros privados, aprobado por Real Decreto Legislativo 6/2004, de 29 de octubre, así como los fondos de pensiones y sus entidades gestoras.
- Cuando durante un ejercicio social hubiesen recibido subvenciones o ayudas con cargo a los presupuestos de las Administraciones Públicas o a fondos de la

⁵ <https://www.boe.es/eli/es/l/2015/07/20/22/con>

	Programa Superior en Big Data & Business Analytics 7ªEd., Sevilla, Marzo – Septiembre 2020	
Proyecto: Herramienta de predicción de auditorías para empresas		
Autores: ANR – JALF – JMR – MCP – MER – RNM		Versión: 2.0

Unión Europea por un importe total acumulado superior a 600.000 euros (auditoría obligatoria de las cuentas anuales correspondientes a dicho ejercicio y a los ejercicios en que se realicen las operaciones o ejecuten las inversiones correspondientes).

- Cuando durante un ejercicio económico hubiesen celebrado con el Sector Público contratos por un importe total acumulado superior a 600.000 euros, y éste represente más del 50 % del importe neto de su cifra anual de negocios (auditoría obligatoria de las cuentas anuales correspondientes a dicho ejercicio y las del siguiente a éste).
- Las sociedades mercantiles que durante dos ejercicios consecutivos reúnan, a la fecha de cierre de cada uno de ellos, al menos dos de las circunstancias siguientes:
 - Que el total de las partidas del activo supere 2.850.000 euros.
 - Que el importe neto de su cifra anual de negocios supere 5.700.000 euros.
 - Que el número medio de trabajadores empleados durante el ejercicio sea superior a 50.
- Las fundaciones en las que, a fecha de cierre del ejercicio, concurren al menos dos de las circunstancias siguientes:
 - Que el total de las partidas del activo supere 2.400.000 euros.
 - Que el importe neto de su volumen anual de ingresos por la actividad propia más, en su caso, el de la cifra de negocios de su actividad mercantil sea superior a 2.400.000 euros.
 - Que el número medio de trabajadores empleados durante el ejercicio sea superior a 50.
- Las entidades en las que se incluyan en sus estatutos la obligación de auditar o cuando así lo acuerden los socios en junta general.

	Programa Superior en Big Data & Business Analytics 7ªEd., Sevilla, Marzo – Septiembre 2020	
Proyecto: Herramienta de predicción de auditorías para empresas		
Autores: ANR – JALF – JMR – MCP – MER – RNM		Versión: 2.0

- Cuando los socios que representen el 5% o más del capital social lo soliciten al registrador mercantil del domicilio social (siempre que no hubieran transcurrido tres meses a contar desde la fecha de cierre de dicho ejercicio).

	Programa Superior en Big Data & Business Analytics 7ªEd., Sevilla, Marzo – Septiembre 2020	
Proyecto: Herramienta de predicción de auditorías para empresas		
Autores: ANR – JALF – JMR – MCP – MER – RNM		Versión: 2.0

3. Modelo de negocio

3.1 Canvas

La Figura 5 muestra el Canvas del modelo de negocio de MLA², la que se identifican los siguientes componentes:

- **Socios clave:** Proveedores de datos (eInforma e Infocif principalmente), consultores específicos y proveedores de tecnología (Microsoft Azure).
- **Actividades clave:** Diseño del modelo predictivo y de una interfaz de usuario simple (dashboard Power BI).
- **Recursos clave:** Infraestructura tecnológica y proveedor de datos. Recursos humanos (ML, científico datos y comercial).
- **Propuesta de valor:** Dotar de información personalizada a auditores para obtener anticipación en el mercado frente a posibles competidores, así como de una wiki de empresas con información de estas. Aumentar la cartera de potenciales clientes de una sociedad auditora.
- **Relaciones con clientes:** Recurrente/Premium con servicios añadidos (búsqueda ampliada geográfica, amplitud de diversidad de los datos que ofrece el dashboard...)
- **Canales:** Foros y Congresos organizados por Colegio de Economistas, ICAC o cualquier otro organismo provincial/autonómico relacionado con sector auditoría, LinkedIn y E-mail Marketing.
- **Segmentos de clientes:** Ámbito nacional. Auditor provincial/local, no *Big Four*.
- **Estructura de costes:** Infraestructura tecnológica, Recursos humanos, Desarrollo del negocio, gastos de operación y gestión de la compañía.
- **Fuente de ingresos:** Servicio recurrente (tarifa mensual), Escalabilidad futura del negocio (añadir valor a los datos).

Proyecto: Herramienta de predicción de auditorías para empresas

Autores: ANR – JALF – JMR – MCP – MER – RNM

Versión: 2.0



Figura 5: Canvas del modelo de negocio

	Programa Superior en Big Data & Business Analytics 7ªEd., Sevilla, Marzo – Septiembre 2020	
Proyecto: Herramienta de predicción de auditorías para empresas		
Autores: ANR – JALF – JMR – MCP – MER – RNM		Versión: 2.0

3.2 Caso de uso

Dada la conexión del equipo del proyecto MLA² con el mundo de la auditoría, se podrían obtener los primeros casos de uso con algunos auditores de la zona de Sevilla, de esta forma sería posible analizar, en el primer mes de implantación, posibles errores o soluciones a tener en cuenta de cara al futuro.

Una vez se obtenga el producto mínimo viable, el primer movimiento será acudir a los foros y congresos que se establecen, mediante varios medios como los colegios de economistas de las principales provincias españolas, el Instituto de Contabilidad y Auditores de Cuentas (ICAC) o el Instituto de Censores Jurados de Cuentas de España (ICJCE). En estos congresos se daría a conocer el servicio de MLA². Al mismo tiempo se precisaría de una fuerte actuación comercial que del mismo modo fuese contundente y rápida.

El servicio de MLA² está enfocado a un objetivo de auditores locales/provinciales, ya sean firmas o auditores independientes, que tengan dificultades a la hora de encontrar clientes en el mercado o de moverse por distintas zonas geográficas o sectores mercantiles diferentes.

El objetivo principal del servicio que MLA² proporcionará a sus futuros clientes es la democratización del sector de la auditoría de cuentas, proporcionando información privilegiada y exclusiva, que en ningún caso la competencia (dentro del sector de auditoría de cuentas) puede obtener con la fiabilidad y rapidez con que MLA² puede proporcionarla. Gracias a ello, estará en las propias manos de cada cliente, la eficacia de su actuación comercial y, por tanto, una probabilidad más equiparada de obtener ingresos.

El valor añadido de MLA² es el proporcionar una información de futuro privilegiada, personalizada y enfocada al negocio. Será la herramienta principal a la que acudir a la hora de tomar decisiones empresariales, tales como invertir en el acercamiento comercial y/o marketing dirigido a determinados sectores empresariales o clientes determinados; planificación anual de auditorías y del equipo humano necesario; entre otras.

Al mismo tiempo, la herramienta proporcionada por MLA² se convertirá en la herramienta básica de trabajo de campo. Esto será posible mediante una simple URL conectada a Power BI, el cual es accesible, económico, y con un uso en alza en el presente de la auditoría de cuentas. Power BI permite también la exportación de datos

	Programa Superior en Big Data & Business Analytics 7ªEd., Sevilla, Marzo – Septiembre 2020	
Proyecto: Herramienta de predicción de auditorías para empresas		
Autores: ANR – JALF – JMR – MCP – MER – RNM		Versión: 2.0

a Excel, herramienta base por excelencia del trabajo de la auditoría de cuentas. Esta herramienta ofrecerá datos recurrentes de las cuentas anuales, que los auditores necesitan en el día a día. Este servicio, adicional a la información que MLA² proporciona en exclusiva para la previsión de nuevos clientes, será clave para la fidelización de clientes y un punto muy importante a tener en cuenta en el desarrollo de MLA² como startup.

Como toda empresa, el fin principal de obtención de ingresos se basará en la captación de clientes y su mantenimiento en el tiempo. MLA² ha determinado para ello unos precios situados en 3 rangos, que se basarán en el volumen de información proporcionada, el uso y la fidelización con el producto/servicio.

Estos paquetes están fijados tal como se describen a continuación:

	Estándar	Plus	Premium
N.º de localizaciones	1	2	5
Método de localización	Código postal	Código postal	Código postal
Límite *	3	6	15
VisualMLA² **	Incluido	Incluido	Incluido
Tarifa Mensual	50,00 euros	70,00 euros	135,00 euros
Extra de Exclusividad ***	150,00 euros	300,00 euros	50 euros por empresa
Tarifa Anual (previsión)	600,00 euros	840,00 euros	1.620,00 euros
Tarifa Anual + Extra (previsión)	750,00 euros	1.140,00 euros	2.370,00 euros

* 3 Empresas ordenadas por proximidad a cada localización.

** Power BI de visualización.

*** En paquetes estándar y plus modalidad pago único por total de empresas exclusivas.

Tabla 1: Paquetes de comercialización del servicio

El precio inicial en el paquete estándar será 50 euros mensuales. Éste conllevará un límite de una sola búsqueda, en la localización que el cliente determine. Esta localización dará a conocer al cliente 3 empresas auditables en el ejercicio correspondiente, por proximidad al punto geográfico determinado.

Las Tarifas Plus y Premium conllevarán un incremento en el número de localizaciones y, por consiguiente, un mayor número de empresas auditables.

La diferencia principal de la Tarifa Premium será la posibilidad de determinar el número de empresas por localización, en lugar de tener el número fijo 3 por cada una. Por

	Programa Superior en Big Data & Business Analytics 7ªEd., Sevilla, Marzo – Septiembre 2020	
Proyecto: Herramienta de predicción de auditorías para empresas		
Autores: ANR – JALF – JMR – MCP – MER – RNM		Versión: 2.0

ejemplo, realizando la búsqueda de las 5 localizaciones demandando 5 empresas para una sola localización y 2 empresas en cada una de las otras cuatro localizaciones.

Asimismo, la información de empresas auditables localizadas en un punto seleccionado por el cliente y que se solicite en exclusiva (Extra de exclusividad - información privilegiada y no compartida con otros auditores clientes de MLA²), tiene un precio estándar de 50 euros por empresa. Este precio se ha fijado sin descuentos por volumen, debido a que es el punto más fuerte del proyecto. Una información privilegiada a tiempo real que no se comparte y totalmente exclusiva para nuestro cliente.

También en el caso de la Tarifa Premium, existirá una diferenciación respecto a las demás tarifas. El cliente podrá decidir si prefiere todas las empresas auditables (15) como extras de exclusividad o por el contrario solo algunas al precio de 50 euros por empresa auditable. En los casos de las Tarifas Estándar y Plus, solo se aceptará el pago único de las 3 empresas auditables en exclusiva por localización.

Por último, se añade en todas las tarifas y de manera gratuita la herramienta de visualización de datos de las empresas auditables, a la que denominamos VisualMLA². El comentado Power BI se incorporará en los equipos del cliente auditor mediante una simple URL. Este servicio será gratuito en la fase inicial de MLA² y su objetivo es la fidelización del cliente auditor. En el futuro, pueden existir modificaciones sobre la herramienta de visualización en función de las demandas del cliente y por consiguiente existirá una evolución positiva, con vistas a incrementar el precio de las tarifas y por tanto de los ingresos, una vez se adquiera una cuota de mercado interesante.

	Programa Superior en Big Data & Business Analytics 7ªEd., Sevilla, Marzo – Septiembre 2020	
Proyecto: Herramienta de predicción de auditorías para empresas		
Autores: ANR – JALF – JMR – MCP – MER – RNM		Versión: 2.0

3.3 Mapa estratégico

Objetivos estratégicos

El objetivo es poder presentar a un usuario no acostumbrado al manejo de este tipo de herramientas, ni conocedor de tecnologías *Big Data*, el estado económico de empresas que pueden ser sus potenciales clientes en base a una predicción de la necesidad de ser auditada en un máximo de 12 meses vista.

Se definen a continuación los objetivos estratégicos de MLA²:

- En el primer año:
 - EBITDA del 0%.
 - 300 clientes.
 - Participar en 2 eventos nacionales de auditoría.
 - Publicar un artículo en 2 medios especializados del sector.
- En el segundo año:
 - EBITDA del 10%.
 - Incrementar la cifra de ingresos en un 30%.
 - Participar en todos los eventos provinciales de auditoría que se convoquen en las 10 principales provincias de España por número de empresas auditadas.
 - Continuar participando en los eventos nacionales de auditoría de años anteriores.
 - Duplicar aparición en los medios respecto al año anterior.

	Programa Superior en Big Data & Business Analytics 7ªEd., Sevilla, Marzo – Septiembre 2020	
Proyecto: Herramienta de predicción de auditorías para empresas		
Autores: ANR – JALF – JMR – MCP – MER – RNM		Versión: 2.0

Perspectivas

El modelo de predicción deberá poder explicar el por qué llega a dicho pronóstico, de manera que incluso una persona no versada en algoritmos de *Machine Learning* pueda seguir la razonabilidad de la predicción. Por tanto, se utilizarán preferentemente modelos lineales tipo random-forest o su variante más moderna boosting, que son capaces de dar resultados de buena calidad con pocos datos de partida.

Indicadores

Los indicadores a mostrar estarán centrados alrededor de la probabilidad de que una determinada empresa cumpla los requisitos de auditoría al año siguiente. Por ejemplo:

- Atendiendo al tamaño
 - Activo del balance
 - Cifra de negocios
 - Número medio de trabajadores
- Por su relación con las AAPP
 - Si ha recibido subvenciones > 600 k euros
 - Si ha contratado > 600 k euros con cargo a las AAPP, siempre que supere el 50% del INCN.

Dashboards

La información se mostrará en un *dashboard* en el que se podrá filtrar y navegar dinámicamente por:

- Localización (CA, provincia, etc)
- Tamaño de la empresa
- Sector (CNAE)

	Programa Superior en Big Data & Business Analytics 7ªEd., Sevilla, Marzo – Septiembre 2020	
Proyecto: Herramienta de predicción de auditorías para empresas		
Autores: ANR – JALF – JMR – MCP – MER – RNM		Versión: 2.0

- Probabilidad resultante del modelo
- Además de los parámetros de base para los indicadores anteriores

	Programa Superior en Big Data & Business Analytics 7ªEd., Sevilla, Marzo – Septiembre 2020	
Proyecto: Herramienta de predicción de auditorías para empresas		
Autores: ANR – JALF – JMR – MCP – MER – RNM		Versión: 2.0

3.4 Beneficios esperados

El modelo de negocio de MLA² no requiere de una inversión variable en su primer ejercicio, aunque sí tiene una estructura de costes bastante fija y alta.

Esto reportará importantes beneficios en un futuro optimista en el que la cuota de mercado sea amplia, aproximadamente con 500 clientes y MLA² se convierta en referencia en el sector de la auditoría en España.

El principal objetivo es fidelizar clientes. Por ello, la estructura de precios e ingresos de MLA² se ha fijado mediante una tarifa mensual, tal y como se describen en el caso de uso, los cuales tienen unos precios competitivos y accesibles, que se elevan al subir de rango de tarifa, pero que a su vez llevan un descuento proporcional al uso de nuestra herramienta (mayor cantidad de localizaciones y por tanto de información).

De cara al futuro, la idea es fidelizar a los clientes con la herramienta de visualización, adaptada para cada cliente, gracias a la escalabilidad del proyecto de MLA².

	Programa Superior en Big Data & Business Analytics 7ªEd., Sevilla, Marzo – Septiembre 2020	
Proyecto: Herramienta de predicción de auditorías para empresas		
Autores: ANR – JALF – JMR – MCP – MER – RNM		Versión: 2.0

4. Planes operativos

4.1 Plan de implantación y distribución

Tras valorar las alternativas para la comercialización del servicio y, a la vista de los casos de uso previamente comentados, se ha optado por un modelo de negocio por suscripción. Al valorar los riesgos, así como los costes, es lo que mejor se adecua a nuestra idea de negocio.

En la suscripción habrá 3 tarifas o paquetes, que estarán basados en el volumen de información y el uso del producto/servicio, como se indica en la Tabla 1: Paquetes de comercialización del servicio.

La suscripción es por un año, con la idea de que disuadir a un potencial cliente de intentar hacer uso de la información y posteriormente darse de baja. El precio inicial del paquete comienza en un mínimo de 600 euros anuales (facturables en cuotas mensuales), que podrá incrementar en base a los diferentes niveles de servicio (Estándar, Plus y Premium).

La diferencia entre las tarifas conlleva límites en las búsquedas de la localización de empresas basadas en los requerimientos de nuestros clientes. Según el paquete el cliente podrá obtener 3, 5 o 15 empresas auditables próximas al punto geográfico que previamente ha determinado el auditor.

El auditor tendrá también la opción de solicitar exclusividad en su producto, es decir que la información sea privilegiada y no se comparta con otro cliente de MLA². Esta opción tiene un precio estándar de 50 euros por empresa exclusiva, que en el caso de las tarifas Estándar y Plus deben ser para el total de las empresas.

Por otro lado, además de la información sobre futuras empresas auditables se les ofrece de manera gratuita la herramienta de visualización de datos de estas empresas con el objetivo de fidelizar al auditor como cliente de MLA².

Con esta herramienta los auditores que tengan acceso a ella, es decir socios y jefes de equipo podrán acceder a datos de la cuenta de pérdidas y ganancias, así como balance, número de empleados... de varios ejercicios, para que puedan realizar una revisión analítica y un estudio de su cliente potencial con el objetivo de tener una visión de su situación, viendo dónde podrían existir complicaciones en el proceso de auditoría, es

	Programa Superior en Big Data & Business Analytics 7ª Ed., Sevilla, Marzo – Septiembre 2020	
Proyecto: Herramienta de predicción de auditorías para empresas		
Autores: ANR – JALF – JMR – MCP – MER – RNM		

decir identificando las áreas críticas de las empresas. Esto ayudará al auditor a enfocar el proyecto, previendo las necesidades de horas del proyecto, así como el equipo que necesitaría para sacar el informe. Disminuiría el tiempo invertido en el estudio del cliente y la propuesta de auditoría estaría más ajustada en precio para su potencial cliente.

En todo momento MLA² contaría con la ayuda de un comercial, este se encargaría de tratar con auditoras, analizando con ellas sus necesidades y explicando los distintos paquetes que se ofrecen, así como las utilidades de la herramienta. Se encargaría también de ayudar ante cualquier dificultad conectándoles y recomendándole a los informáticos cualquier mejora o recomendación.

El comercial es el encargado por tanto de vender nuestro producto, nuestro principal fuerte en la implantación y distribución, la cara de MLA².

Para ello junto con el plan de marketing detallado en los siguientes puntos, MLA² cuenta con unos gastos de representación. El comercial será realmente en encargado de encuadrar y organizar los eventos a los que asistirá. Estos gastos conllevan el viaje, el pago de eventos y patrocinios.

Su trabajo por tanto encuadrará tanto la búsqueda de futuros clientes, así como preparación para los eventos y búsqueda de nuevas formas de marketing. Posteriormente tendrá que encargarse de contactar con los clientes teniendo en cuenta que en todo momento se ofrece un servicio personalizado y para el que se requiere una buena comunicación con los auditores.

Por otro lado, el comercial será, como se ha comentado, el punto de conexión entre el equipo técnico y el cliente, por lo que debe de conocer de primera mano la herramienta incluso siendo capaz de solucionar algunas de las cuestiones que les planteen los clientes. Recogiendo recomendaciones que se plantearan para la futura evolución y expansión.

El comercial tendrá una retribución fija mensual a la que además se añadirá una comisión por clientes obtenidos. En el plan económico financiero se considera la retribución del comercial, que percibirá un salario fijo anual de 30.000 euros, junto con una variable de 10.000 euros en base a objetivos revisables año tras año.

	Programa Superior en Big Data & Business Analytics 7ªEd., Sevilla, Marzo – Septiembre 2020	
Proyecto: Herramienta de predicción de auditorías para empresas		
Autores: ANR – JALF – JMR – MCP – MER – RNM		Versión: 2.0

4.2 Plan de expansión

Una vez el proyecto sea conocido, al ser una herramienta muy general y donde la contabilidad es la base, se podría aplicar en otros países revisando únicamente sus requisitos de auditabilidad.

Además, nos planteamos desarrollar nuevos servicios en la herramienta de visualización en función de las demandas que vaya requiriendo el cliente, sirviendo de base de datos y dando la posibilidad de nuevos usos como podría ser circularizaciones, o selecciones de proveedores y clientes con creando una herramienta de selección aleatoria. De las demandas y sugerencias se encarga de comunicarlas y plantearlas el Comercial.

Por otro lado, MLA² busca una expansión diferente, donde las empresas puedan anunciarse en su herramienta en búsqueda de un auditor, es denominado *Marketplace*. Hablamos de empresas que sin tener la obligación de auditarse requieran auditar sus cuentas de forma voluntaria. De esta forma necesitarían buscar auditores, teniendo con el *Marketplace* la opción de anunciar su necesidad de auditoría, las auditoras que tengan contratados los servicios MLA² recibirán estas notificaciones pudiendo ponerse en contacto con el anunciante enviándole su propuesta.

	Programa Superior en Big Data & Business Analytics 7ª Ed., Sevilla, Marzo – Septiembre 2020	
Proyecto: Herramienta de predicción de auditorías para empresas		
Autores: ANR – JALF – JMR – MCP – MER – RNM		Versión: 2.0

4.3 Plan de marketing

Tras un análisis de las opciones disponibles para dar a MLA² a conocerse entre su público objetivo, optamos por un marketing tanto online como cara a cara (offline). (Apartado 3.1)

Marketing Online: se creará una Web de empresa siendo esta nuestra primera toma de contacto con el futuro cliente que nos conoce de oídas y que quiere saber más sobre nuestro servicio buscando información sobre MLA². La presencia online es fundamental en nuestro negocio. A esto le uniríamos campañas por email para tener comunicación con los clientes para ofertas o nueva información de nuestros productos, así como invitaciones a eventos en los que participemos.

También se implantará un plan de marketing digital en *LinkedIn*.

En relación con lo anterior estaría nuestro Marketing offline, acudiríamos a los encuentros que organiza el Instituto de Auditores Internos, estos encuentros tienen el objetivo de ofrecer colaboración y *networking* entre las auditoras, así como evaluar las novedades en la profesión.

El instituto de Auditores nos proporciona los lunes de Instituto, foros de expertos, así como Jornadas anuales.

Las Jornadas anuales serían nuestra principal exposición a los clientes donde ponentes y asistentes comparten experiencias y se dan a conocer herramientas para aplicar en su día a día y en su visión del futuro.

Por otro lado, un lunes al mes el Instituto organiza reuniones que tratan temas de tendencias y presentan herramientas y estudios, por lo que sería la principal forma de dar a conocer la empresa en un ambiente en el que además los auditores se encuentran cómodos y confiados.

Además del Instituto de auditores contamos con instituto de contabilidad y auditoría de cuentas (ICAC), Colegio de economistas y el Instituto de Censores Jurados de Cuentas de España (ICJCE).

El ICAC se encarga de la formación continua de los auditores con jornadas conferencias y actividades homologadas.

	Programa Superior en Big Data & Business Analytics 7ªEd., Sevilla, Marzo – Septiembre 2020	
Proyecto: Herramienta de predicción de auditorías para empresas		
Autores: ANR – JALF – JMR – MCP – MER – RNM		Versión: 2.0

5. Solución tecnológica

Antes de abordar el detalle de la solución técnica, conviene hacer mención a las tres premisas sobre las que gira este proyecto:

- el objetivo es poder presentar a un usuario no acostumbrado al manejo de este tipo de herramientas, ni conocedor de tecnologías Big Data, el estado económico de empresas que pueden ser sus potenciales clientes en base a una predicción de la necesidad de ser auditada en un máximo de 12 meses vista.
- el modelo de predicción deberá poder explicar el por qué llega a dicho pronóstico, de manera que incluso una persona no versada en algoritmos de *Machine Learning* pueda seguir la razonabilidad de la predicción.
- la visualización de los resultados tendrá pues un peso específico en el producto.

	Programa Superior en Big Data & Business Analytics 7ªEd., Sevilla, Marzo – Septiembre 2020	
Proyecto: Herramienta de predicción de auditorías para empresas		
Autores: ANR – JALF – JMR – MCP – MER – RNM		Versión: 2.0

5.1 Análisis funcional

El siguiente diagrama conceptual indica las fases principales del proceso, desde la captura de información de diferentes proveedores hasta la visualización vía internet por el usuario final.

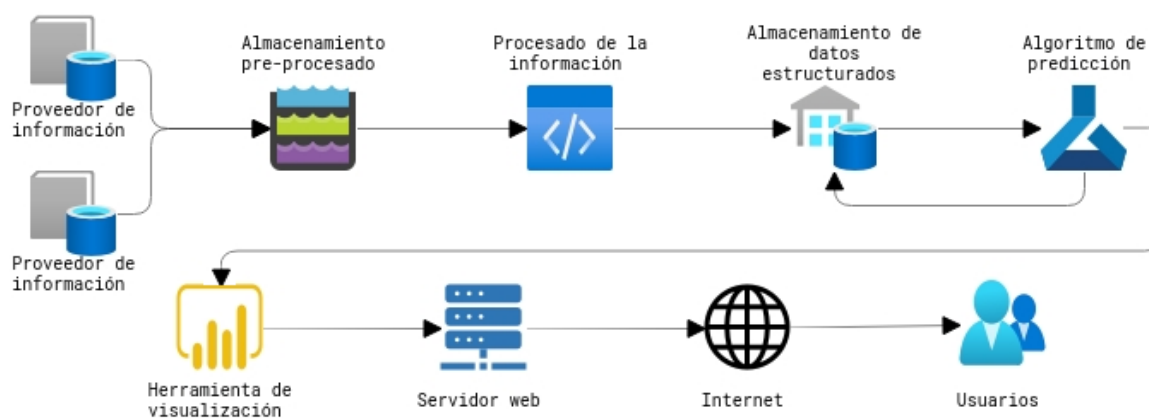


Figura 6: Diagrama funcional del flujo de procesamiento de datos

- **Proveedores:** se incluyen aquí proveedores de datos financieros como elInforma, Infocif y Axesor entre otros, así como el Registro mercantil, datos macroeconómicos, bases de datos de empresas auditadas, etc. Se dará prioridad a aquellos proveedores que dispongan de API pero, en caso de no tenerla, la información se extraerá mediante *scraping* de la web, archivos de intercambio, etc. Dado que la información necesaria de cada empresa para alimentar el modelo no varía dentro del año fiscal salvo en casos muy excepcionales, consideraremos que los datos son estáticos (*data at rest*)
- **Almacenamiento pre-procesado:** en una base de datos NoSQL se almacenarán todos los datos capturados de todas las fuentes, para convertirlos mediante un procesamiento posterior en un conjunto de datos estructurados que alimenten el modelo de predicción
- **Procesado de la información:** en esta fase se procede a la conversión de datos no estructurados en datos estructurados que puedan ser digeridos por el modelo de predicción y por la herramienta de visualización del cuadro de mando/*dashboard*.

	Programa Superior en Big Data & Business Analytics 7ªEd., Sevilla, Marzo – Septiembre 2020	
Proyecto: Herramienta de predicción de auditorías para empresas		
Autores: ANR – JALF – JMR – MCP – MER – RNM		Versión: 2.0

- **Almacenamiento estructurado:** se trata de una base de datos SQL en la que se almacenará el resultado del procesamiento anterior junto con los resultados de la predicción del modelo predictivo.
- **Algoritmo de predicción:** éste es el corazón del proceso que, mediante un algoritmo de ML, realizará la predicción de las empresas que serán susceptibles de ser auditadas en el próximo ejercicio fiscal.
- **Herramienta de visualización:** esta es la herramienta de presentación de los resultados del algoritmo de predicción al usuario final y toda la información útil acerca de la empresa para poder realizar una auditoría de la misma. Se trataría de una herramienta *cloud* integrada en la web de MLA² que mostraría dichos resultados en forma de KPIs en un *dashboard*, según se detalla en el apartado 5.6.
- **Servidor web:** la herramienta de visualización se publicará dentro de la web de MLA², con acceso restringido a usuarios de pago, de forma que estos puedan consultar la información desde cualquier lugar, previa autenticación.
- **Usuario:** es el usuario final de la herramienta, auditor o persona integrada en un equipo de auditoría.

	Programa Superior en Big Data & Business Analytics 7ªEd., Sevilla, Marzo – Septiembre 2020	
Proyecto: Herramienta de predicción de auditorías para empresas		
Autores: ANR – JALF – JMR – MCP – MER – RNM		Versión: 2.0

5.2 Fuentes de datos

Para obtener el objetivo estratégico expresado en el apartado anterior, serán necesarios obtener para cada empresa los datos de:

- Activos
- Importe neto de la cifra de negocio
- Empleados

que se pueden obtener (en algunos casos vía API, aunque de manera muy limitada) en:

- Cámara de Comercio de España: <https://www.camerdata.es/base-de-datos-empresas-espanolas>
- API para el Registro Mercantil: <https://librebor.me/>
- API de IberInform: <https://www.iberinform.es/productos/api-integracion>
- eInforma: <https://www.einforma.com/>
- Infocif: <https://www.infocif.es/>

Cuantas más empresas alimenten el modelo, más preciso será éste. Sin embargo, como veremos en el apartado 6 a propósito del análisis económico del proyecto, los datos de detalle son caros cuando se pretenden conseguir para la totalidad de las empresas españolas. Considerando además que nuestro *target* de cliente son auditores de empresas pyme, se hace necesario un filtrado previo de las más de 3.36 millones de empresas que hay en nuestro país.

Para ello, utilizaremos la base de datos de

<https://www.centraldecomunicacion.es/ofertas-base-datos-empresas-espana/>

que facilita la siguiente información de las 840.000 empresas más importantes y que además están activas, eliminando de facto las que con un alto grado de probabilidad no serán auditables en el medio plazo o que sean empresas zombis o en vías de cierre:

	Programa Superior en Big Data & Business Analytics 7ª Ed., Sevilla, Marzo – Septiembre 2020	
Proyecto: Herramienta de predicción de auditorías para empresas		
Autores: ANR – JALF – JMR – MCP – MER – RNM		Versión: 2.0

- Facturación anual
- CIF de la empresa
- CNAE registrado
- Población
- Sitio web
- Teléfono
- Marcas Registradas por cada empresa
- Dirección
- Localidad
- Código Postal

Además, el modelo se alimentará de información pública suministrada por el Instituto Nacional de Estadística (INE) con datos macroeconómicos desagregados al mayor nivel geográfico disponible (provincial o autonómico, según el caso) y que pueden ayudar a mejorar la predicción, así como el impacto de un entorno macroeconómico volátil:

- Tasa de empleo
- Tasa de paro
- Tasa de actividad
- IPC (en variación interanual porcentual)
- PIB absoluto (en euros y variación porcentual)
- PIB per cápita (en euros y variación porcentual)

Por último, se usará una fuente de datos pública con las coordenadas de los códigos postales españoles, y que permitirá visualizar las empresas en un mapa en el cuadro de mando, calcular distancias, etc

 Escuela de organización industrial	Programa Superior en Big Data & Business Analytics 7ªEd., Sevilla, Marzo – Septiembre 2020	
Proyecto: Herramienta de predicción de auditorías para empresas		
Autores: ANR – JALF – JMR – MCP – MER – RNM		Versión: 2.0

5.3 Ingesta de datos

Partiendo de la base de datos de Central de Comunicación (CdC) identificada en el apartado anterior, se realizará un filtrado para optimizar los costes de adquisición de los datos, tomando en cuenta solo el criterio de facturación. Para ello, consideraremos sólo las que lo cumplen en el último año con cuentas publicadas, de manera que el año en curso sería el que, de cumplirse también, haría que la auditoría contable deviniese en obligatoria si concurre al menos uno de los otros requisitos.

En base a las empresas que facturan más de 5.7 Meuros, obtendremos una tabla como la siguiente con, aproximadamente, 80.000 registros (dependerá del año) con la facturación de cada empresa en el último año, aunque lo que interesa al proceso de tratamiento de datos no es la facturación sino el CIF de esas empresas, para las que en una etapa posterior se adquirirán datos de detalle:

TABLA CdC	
CIF	VentasCdC_X-1 (> 5.7 Meuros)

Tabla 2: Tabla CdC

Cuyo diccionario de datos es:

- **CIF:** CIF de la empresa
- **VentasCdC_X-1:** importe neto de la cifra de negocio correspondiente al último periodo contable cerrado, según los datos de Central de Comunicación

Como se verá en el apartado 5.8 Prototipo y validación, el prototipo entrenado muestra que el peso de las variables asociadas al Año -3 es bastante bajo, por lo que es muy razonable admitir que el peso de variables asociadas a años aún más anteriores en el tiempo será prácticamente despreciable. Por tanto, y al objeto de reducir no solo un posible *overfitting* sino los costes de adquisición de los datos, el modelo de predicción sólo considerará los datos de los tres últimos años, aunque cada año con diferente propósito.

Para entender mejor el proceso que sigue, conviene tener en mente la siguiente secuencia temporal y para qué se usa la información de cada año. Se hace notar que la pandemia del Covid-19 ha traído consigo una extensión de los plazos para la

	Programa Superior en Big Data & Business Analytics 7ªEd., Sevilla, Marzo – Septiembre 2020	
Proyecto: Herramienta de predicción de auditorías para empresas		
Autores: ANR – JALF – JMR – MCP – MER – RNM		Versión: 2.0

presentación de las cuentas anuales. Es por ello que muchas empresas no han presentado aún los datos de sus cuentas del año 2019 y por lo que se han usado datos del periodo 2016-2018:

Proyecto: Herramienta de predicción de auditorías para empresas

Autores: ANR – JALF – JMR – MCP – MER – RNM

Versión: 2.0

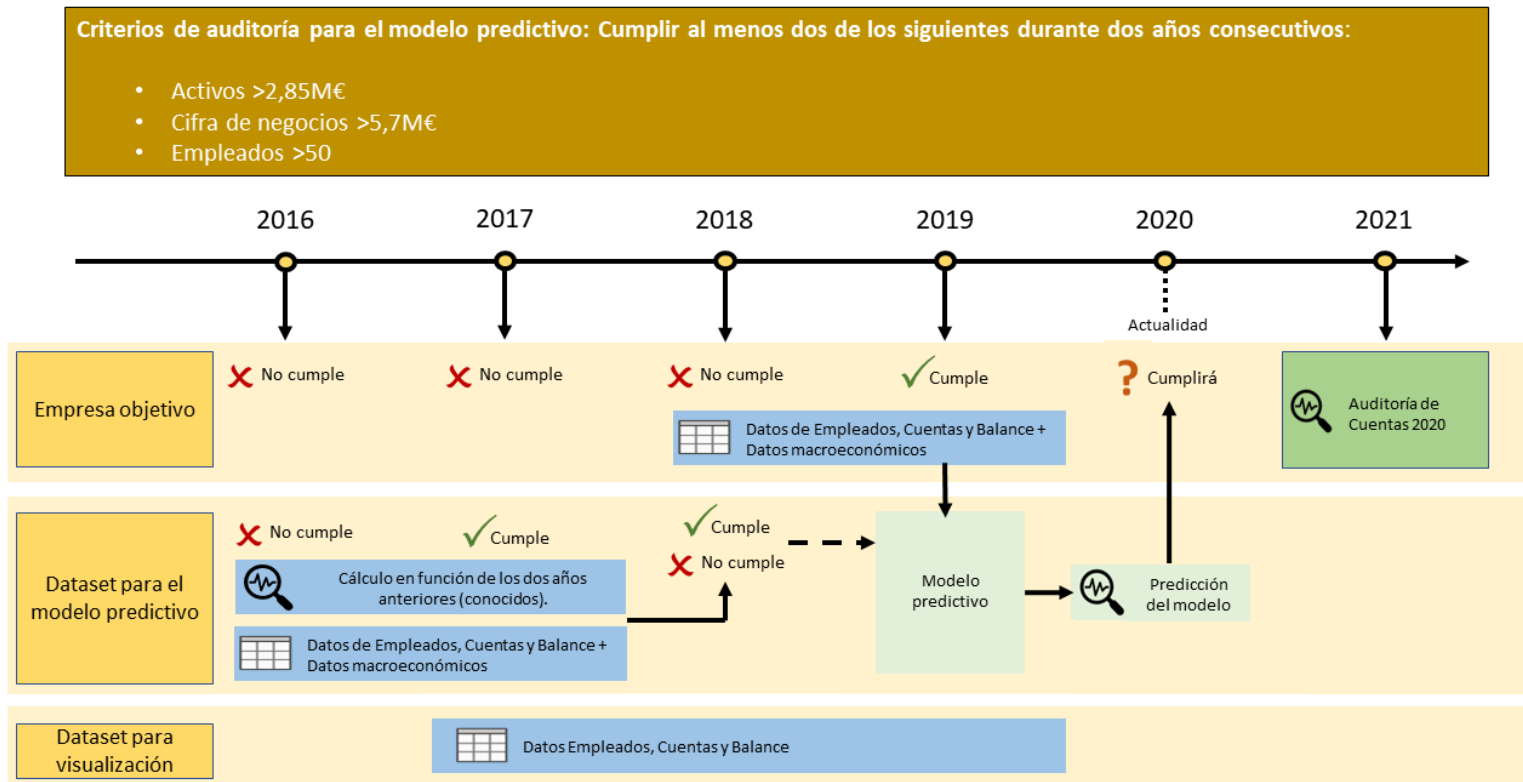


Figura 7: Secuencia temporal del uso de los datos en el modelo predictivo y en la visualización

	Programa Superior en Big Data & Business Analytics 7ªEd., Sevilla, Marzo – Septiembre 2020	
Proyecto: Herramienta de predicción de auditorías para empresas		
Autores: ANR – JALF – JMR – MCP – MER – RNM		Versión: 2.0

Dadas las limitaciones de la APIs desarrolladas por posibles proveedores como elInforma o el Registro Mercantil, finalmente se ha decidido que la forma más efectiva y económica de recabar la información de detalle es solicitar a las empresas proveedoras datos (elInforma, Infocif, Cámara de Comercio, etc), dicha información para cada una de las empresas de la tabla CdC (Tabla 2Tabla 2: Tabla CdC), conforme al siguiente diccionario de datos:

- **Datos generales de la empresa**

Campo	Descripción
Nombre	Nombre legal de la empresa
CIF	CIF de la empresa
Dirección	Dirección de su sede social (SS)
CP	Código Postal de su SS
Municipio	Municipio de su SS
Provincia	Provincia de su SS
CCAA	Comunidad Autónoma de su SS
Teléfono	Teléfono de su SS
Sector	CNAE de la empresa
CargoDirectivo	Nombre del máximo directivo o, en su defecto, de su CFO
Empleados	Nº de empleados de la empresa

*Tabla 3: Diccionario de la tabla de información microeconómica de las empresas
Campos generales*

- **Datos del balance⁶**

Campo ⁷	Descripción
ActivoNoCorriente	Activo no corriente
InmovilizadoIntangible	Inmovilizado intangible
InmovilizadoMaterial	Inmovilizado material
InversionesInmbiliarias	Inversiones inmobiliarias
InversionesGrupoLargo	Inversiones del grupo a LP
InversionesFinanLargo	Inversiones financieras a LP

⁶ LP: Largo plazo
CP: Corto plazo

⁷ De cara a la integración posterior en una única tabla, como se verá en el apartado 5.4, al nombre de los campos se le añadirá el sufijo 1, 2 o 3, dependiendo si se trata de datos del año -1, -2 o -3

Proyecto: Herramienta de predicción de auditorías para empresas

Autores: ANR – JALF – JMR – MCP – MER – RNM

Versión: 2.0

Campo⁷	Descripción
OtrosActivosLargo	Otros activos a LP
ActivoCorriente	Activo corriente
Existencias	Existencias de inventario
DeudoresComerciales	Deudores comerciales
InversionesGrupoCorto	Inversiones del grupo a CP
InversionesFinancCorto	Inversiones financieras a CP
OtrosActivosCorto	Otros activos a CP
EfectivosActivosLiquidos	Efectivos y activos líquidos
TotalActivo	Activo total
PatrimonioNeto	Patrimonio neto
FondosPropios	Fondos propios
Capital	Capital
Reservas	Reservas
ResultadoEjerc	Resultado del ejercicio
OtrosFondos	Otros fondos
AjustesCambio	Ajustes de cambio
Subvenciones	Subvenciones
PasivoNoCorriente	Pasivo no corriente
ProvisionesLargo	Provisiones a LP
DeudasLargo	Deudas a LP
DeudasGrupoLargo	Deudas del grupo a LP
OtrosPasivosLarog	Otros pasivos a LP
PasivoCorriente	Pasivo corriente
ProvisionesCorto	Provisiones a CP
DeudasCorto	Deudas a CP
DeudasGrupoCorto	Deudas del grupo a CP
AcreedoresComerc	Acreedores comerciales
OtrosPasivos	Otros pasivos
TotalPasivoPatrim	Pasivo Patrimonial Total

*Tabla 4: Diccionario de la tabla de información microeconómica de las empresas
Campos del Balance anual*

Proyecto: Herramienta de predicción de auditorías para empresas

Autores: ANR – JALF – JMR – MCP – MER – RNM

Versión: 2.0

- **Datos de la cuenta de explotación**

Campo⁸	Descripción
IngresosExp	Ingresos de explotación
ImpNetoCifraNegocio	Importe neto de la cifra de negocio
VariacionExistProd	Variación de existencias producción
TrabajosParaActivo	Trabajos para el activo
Aprovisionamientos	Aprovisionamientos
OtrosIngresosExp	Otros ingresos de explotación
GastosPersonal	Gastos de personal
OtrosGastosExp	Otros gastos de explotación
AmortizacionInm	Amortización del inmovilizado
ExcesoProv	Exceso en proveedores
OtrosResultados	Otros resultados
ResultadosExp	Resultados de explotación
IngresosFinanc	Ingresos financieros
GastosFinac	Gastos financieros
DiferenciasCambio	Diferencias de cambio
OtrosResultFinanc	Otros resultados financieros
ResultadoFinanc	Resultado financiero
ResultadoAntesImp	Resultado antes de impuestos
ImpSobreBeneficios	Impuestos sobre beneficios
ResultadoEjercCuentas	Resultado del ejercicio

*Tabla 5: Diccionario de la tabla de información microeconómica de las empresas
Campos de la Cuenta de explotación anual*

Esta información la suministran en una tabla estructurada por cada año y, como se verá en el apartado 6, la opción más económica y que se ha considerado para el análisis financiero es la de elnforma.

Por otra parte, como se verá en el apartado 5.8 con los resultados del prototipo desarrollado, de esas casi 60 variables por año (120 considerando los 2 años que alimentan el modelo predictivo) apenas bastan 10 de ellas para explicar el 80% de la predicción de auditoría, aunque todas ellas son necesarias para el cuadro de mando en Power BI con los datos de detalle de la empresa.

⁸ De cara a la integración posterior en una única tabla, como se verá en el apartado 5.4, al nombre de los campos se le añadirá el sufijo 1, 2 o 3, dependiendo si se trata de datos del año -1, -2 o -3

	Programa Superior en Big Data & Business Analytics 7ªEd., Sevilla, Marzo – Septiembre 2020	
Proyecto: Herramienta de predicción de auditorías para empresas		
Autores: ANR – JALF – JMR – MCP – MER – RNM		Versión: 2.0

En lo referente a los datos macroeconómicos facilitados por el INE, dado su poco volumen -un dato por variable y provincia (52) en el mejor de los casos, cuando no por Comunidad Autónoma (17) -, se capturarán y procesarán manualmente de la web del INE para construir en excel una tabla que asocie dichos datos macro a cada Código Postal (CP) según corresponda, de manera que dicho CP sea el campo que permita integrar posteriormente la información microeconómica de cada empresa con la macroeconómica del entorno. La tabla siguiente muestra el diccionario de datos de cada tabla de datos macroeconómicos anuales:

Campo⁹	Descripción
CP	2 primeros dígitos de los códigos postales de la provincia
TE	Tasa de empleo provincial
TP	Tasa de paro provincial
TA	Tasa de actividad provincial
IPC	IPC provincial
PIBab	PIB absoluto provincial
PIBCAab	Variación PIB provincial
PIBpc	PIB per cápita, provincial
PIBCApc	Variación PIB per cápita, provincial

Tabla 6: Diccionario de las tablas de datos macroeconómicos

Finalmente, para poder visualizar en el cuadro de mando de Power BI la ubicación geográfica de las empresas con predicción positiva, se usará una tabla pública que proporciona las coordenadas (latitud y longitud) del centro de cada código postal:

Campo	Descripción
codigopostalid	Código postal
lat	Latitud del código postal
lon	Longitud del código postal

Tabla 7: Coordenadas de cada código postal español

⁹ De cara a la integración posterior en una única tabla, como se verá en el apartado 5.4, al nombre de los campos se le añadirá el sufijo 1, 2 o 3, dependiendo si se trata de datos del año -1, -2 o -3

	Programa Superior en Big Data & Business Analytics 7ªEd., Sevilla, Marzo – Septiembre 2020	
Proyecto: Herramienta de predicción de auditorías para empresas		
Autores: ANR – JALF – JMR – MCP – MER – RNM		Versión: 2.0

Además, esta información nos permitirá calcular la distancia entre dos códigos postales a través de la fórmula *haversine*¹⁰, lo que se usará para filtrar los resultados en función de la distancia al código postal seleccionado como base.

¹⁰ La fórmula de *haversine* determina la distancia de círculo máximo entre dos puntos en una esfera dado sus longitudes y latitudes. Más información en https://es.qwe.wiki/wiki/Haversine_formula

	Programa Superior en Big Data & Business Analytics 7ªEd., Sevilla, Marzo – Septiembre 2020	
Proyecto: Herramienta de predicción de auditorías para empresas		
Autores: ANR – JALF – JMR – MCP – MER – RNM		Versión: 2.0

5.4 Enriquecimiento y unificación

La información de las fuentes de datos descritas en el apartado anterior se consolidará en una tabla en base a las siguientes relaciones a partir de la tabla con los datos generales de las empresas:

- Se usará el campo CIF para relacionar los datos microeconómicos de detalle (balance y cuenta de resultados)
- Se usará el campo Código Postal (CP) para relacionar los campos de posicionamiento geográfico para el módulo de visualización y los datos macroeconómicos provinciales o regionales.

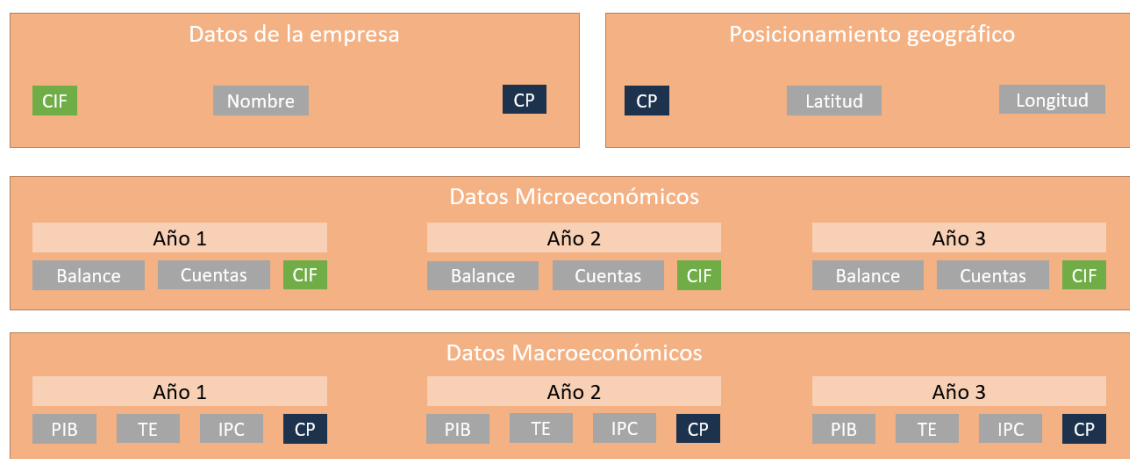


Figura 8: Modelo de datos de la tabla con todos los datos integrados

Estas relaciones darán lugar a una tabla de datos estructurados, con tantos registros como empresas hayan facturado en el último año al menos 5.7 M euros y con más de 200 variables (campos) para cada empresa.

Esta tabla será la base tanto para la creación de los *dataset* del modelo de predicción como la alimentación del cuadro de mando.

	Programa Superior en Big Data & Business Analytics 7ªEd., Sevilla, Marzo – Septiembre 2020	
Proyecto: Herramienta de predicción de auditorías para empresas		
Autores: ANR – JALF – JMR – MCP – MER – RNM		Versión: 2.0

5.5 Modelo de predicción

Como ya se ha comentado al inicio de este apartado 5, dada la tipología del cliente final usuario y la cantidad de datos de los que dispondremos para alimentar el modelo, la preferencia será por modelos lineales (pues permiten explicar por qué se llega a la predicción). De entre todos los modelos lineales, la elección ha sido la de los ensembles de árboles, que presentan las siguientes características:

- es uno de los algoritmos de aprendizaje más certeros que hay disponible. Para un set de datos lo suficientemente grande produce un clasificador muy certero. Incluso cuando no hay muchos datos, proporciona resultados muy razonables.
- corre eficientemente en bases de datos grandes.
- maneja cientos de variables de entrada sin excluir ninguna.
- da estimaciones de qué variables son importantes en la clasificación.
- tiene un método eficaz para estimar datos perdidos y mantener la exactitud cuando una gran proporción de los datos está perdida.
- computa los prototipos que dan información sobre la relación entre las variables y la clasificación.

Para alimentar el modelo, se partirá de la macro-tabla cuyo modelo de datos es el indicado anteriormente en la Figura 8, para obtener un *dataset* ampliado mediante la incorporación de unos campos adicionales para definir la variable objetivo. En concreto:

	Programa Superior en Big Data & Business Analytics 7ªEd., Sevilla, Marzo – Septiembre 2020	
Proyecto: Herramienta de predicción de auditorías para empresas		
Autores: ANR – JALF – JMR – MCP – MER – RNM		Versión: 2.0

Campo	Descripción
CriteriosAño1	Número de criterios que la empresa cumplió en el año -1
CriteriosAño2	Número de criterios que la empresa cumplió en el año -2
CriteriosAño3	Número de criterios que la empresa cumplió en el año -3
Auditable	Variable categórica que valdrá 1 si la empresa cumplió 2 o más criterios en el año -1 Y en el año -2, y que es la que el modelo debe predecir en base a los datos de los años -2 y -3

Tabla 8: Campos adicionales a la macro-tabla del modelo predictivo

Este *dataset* ampliado se dividirá aleatoriamente en dos, uno con el 80% de los registros para entrenamiento del modelo y otro con el 20% restante para su validación.

El objetivo del modelo es predecir el valor de la variable auditable en el año -1 en base a los datos de los años -2 y -3, validando los resultados con los datos conocidos para los años -1 y -2.

Si el modelo resulta exitoso, bastará con introducir los datos de los 2 últimos años cerrados de una empresa (con al menos dos criterios cumplidos para el último año, aunque no se cumpliera ninguno en el anterior) para predecir si esa empresa cumplirá dos criterios al acabar el año en curso y, por tanto, incurrirá en obligación de auditar sus cuentas.

Proyecto: Herramienta de predicción de auditorías para empresas

Autores: ANR – JALF – JMR – MCP – MER – RNM

Versión: 2.0

5.6 Visualización de resultados

La visualización del resultado del modelo se presenta al cliente en el módulo de visualización llamado *VisualMLA*². Este módulo está construido en Power BI a partir de los datos que devuelve el modelo predictivo y teniendo en cuenta el pack que el cliente ha seleccionado, como se menciona en el apartado 3.2.

El módulo de visualización se presenta en dos páginas:

- **Principal:**

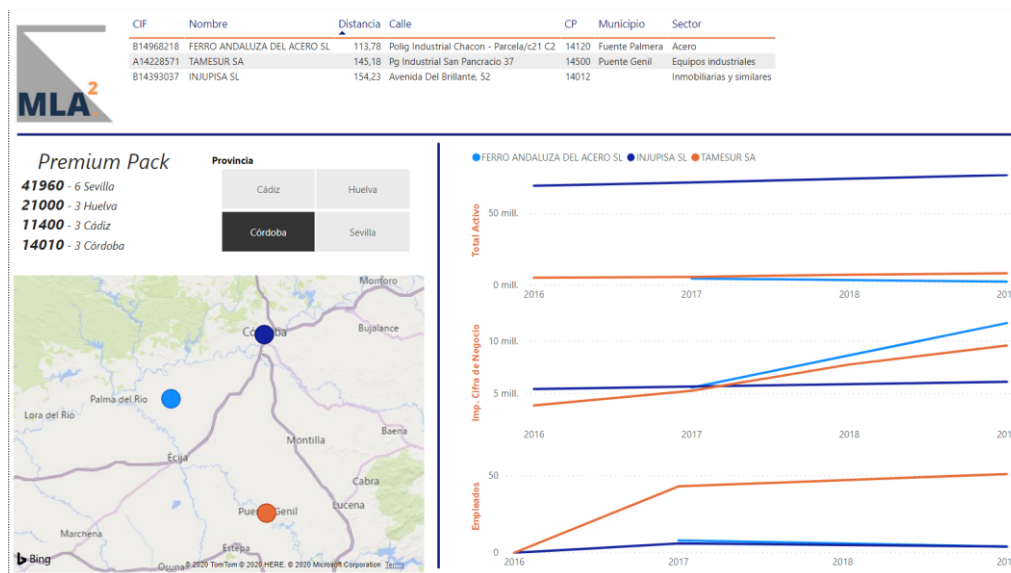


Figura 9: Pantalla inicial Módulo de visualización

En esta hoja se presentan las empresas que devuelve el modelo por código postal en función del pack que tenga el cliente. Ofrece un mapa de localización en función del código postal y una comparativa del Total de Activos, Cifra de Negocio y Empleados.

La idea principal de esta hoja es presentar al cliente de forma muy visual las empresas que puedan resultarle más interesantes en función de la distancia o la magnitud de sus cuentas basadas en el total de activos, la cifra de negocio y los empleados.

● **DetalleEmpresa:**



Figura 10: Pantalla detalle empresa del Módulo de visualización

Esta hoja permite filtrar por cada empresa según su localización y ofrece un desglose de los datos de la empresa, el detalle de su balance y sus cuentas. Se pone especial énfasis visual en las principales agrupaciones contables, de manera que el auditor pueda tener una primera toma de contacto con los resultados de la empresa en el último año y poder así trazar su plan comercial y de acercamiento a esta empresa.

	Programa Superior en Big Data & Business Analytics 7ªEd., Sevilla, Marzo – Septiembre 2020	
Proyecto: Herramienta de predicción de auditorías para empresas		
Autores: ANR – JALF – JMR – MCP – MER – RNM		Versión: 2.0

5.7 Arquitectura de la solución

A partir de las especificaciones de este proyecto se ha optado por utilizar las herramientas de Azure. Esta plataforma es interesante para nuestro proyecto porque dispone de una integración completa con una herramienta de visualización como Power BI, al ser ambas del mismo proveedor.

Siguiendo el modelo conceptual del proyecto descrito en el apartado 5.1 Análisis funcional, se diseña la plataforma de MLA² formada por los componentes descritos en el siguiente esquema:

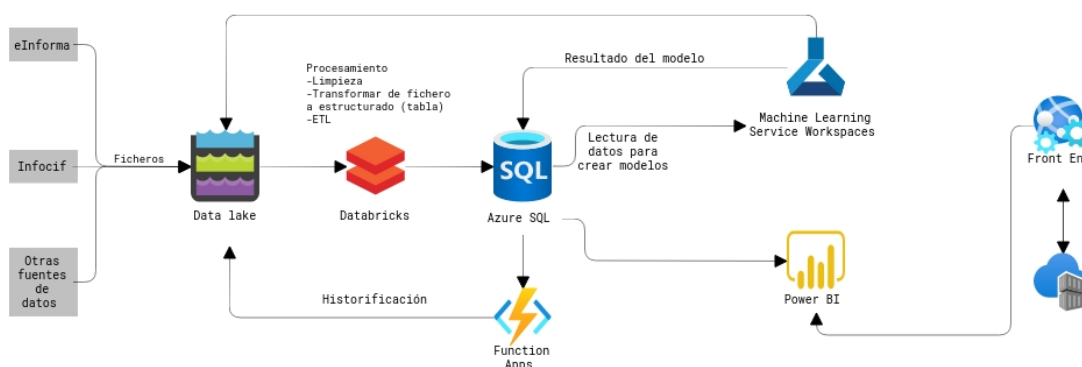


Figura 11: Arquitectura final

- Function Apps:** son instancias de Azure que permiten la ejecución de una función desarrollada en lenguajes como .NET, NodeJS, Java. Este componente se va a aplicar para la recepción de datos de las APIs de las que se van a recibir los informes de las empresas. Además, se va a aplicar este componente para el envío de datos procedente de Power BI al Data Lake donde se almacenará el histórico de la aplicación.
- Data Lake:** es el repositorio en Azure que permite el almacenamiento de una gran cantidad de datos no estructurados, previamente al procesado de esta información procedente de los proveedores. Es donde se guarda la información en primera instancia.
- ETL (Databricks):** se utilizará la herramienta como ETL para la limpieza de datos eliminando posibles errores o incoherencias y transformación de los datos no estructurados en las tablas de datos que se definirán en la base de datos SQL.

	Programa Superior en Big Data & Business Analytics 7ªEd., Sevilla, Marzo – Septiembre 2020	
Proyecto: Herramienta de predicción de auditorías para empresas		
Autores: ANR – JALF – JMR – MCP – MER – RNM		Versión: 2.0

- **SQL:** se utilizará la base de datos de Azure SQL para almacenar los datos estructurados en la plataforma, aprovechando al máximo la integración entre todas las herramientas de Azure.
- **Machine Learning Service Workspaces:** con el objetivo de crear el modelo predictivo se utiliza esta herramienta de Azure que mediante la obtención de los datos procedentes de la base de datos SQL y la aplicación de las funciones *Decision Forest Regression* y *Boosted Decision Tree Regression* ya implementadas en el componente y que permiten de una manera sencilla la creación del modelo.
- **Power BI:** los *dashboards* que permiten al usuario la visualización de los datos de empresas de interés combinados con los resultados del modelo predictivo se mostrarán en un cuadro de mando creado mediante esta herramienta de visualización de Microsoft e integrada en Azure.

	Programa Superior en Big Data & Business Analytics 7ª Ed., Sevilla, Marzo – Septiembre 2020	
Proyecto: Herramienta de predicción de auditorías para empresas		
Autores: ANR – JALF – JMR – MCP – MER – RNM		Versión: 2.0

5.8 Prototipo y validación

Para poder demostrar la validez de la propuesta de MLA², y dado que los datos de las empresas no son públicos de manera gratuita, se ha hecho un proceso de *scraping* a la web de Infocif.

Como se ve en las imágenes siguientes, a través de diversas pantallas, Infocif no sólo ofrece los datos que definen la obligatoriedad de auditar las cuentas (empleados, activo y facturación anual), sino que proporciona un buen desglose del balance y de la cuenta de resultados, lo que permitirá identificar qué componentes de ambos son los mejores predictores de la obligatoriedad de auditar las cuentas.

Información de la compañía		
CIF	B411452855	Sector: Equipos y servicios comerciales
Antigüedad	-	Nº de empleados: 52
Domicilio	Pol Ind La Red, P 35, ctra Sev-mal Km 5,2 41500 - (Alcala De Guadaira) - Sevilla	Cargos directivos - Administrador Mancomunado: LOZANO JIMENEZ JUAN Ver más
Teléfono	955634500	Matriz: -
Registro	Registro Mercantil de Sevilla	
Web	www.gruaslozano.com	

Cuentas anuales (en €)		
	2018	2017
Ingresos	5.779.209	5.112.650
EBITDA	782.626	272.423
Resultado de explotación	756.583	242.328
Total activo	6.743.960	5.719.438
Patrimonio neto	2.509.204	1.349.790
Deuda total	476.043	290.116
Clientes	1.965.175	1.697.377
Proveedores	1.088.300	965.782
Ultimo año cuentas		2018
Auditor	RUBIO MEMBRIVE Y GISBERT SOCIEDAD LIMITADA PROFESI	

Valoración y evaluación crediticia (en €)				
Valoración	7.826.263 10x EBITDA	2.509.204 Valor neto contable		
			2018	2017
Ratios Deuda total / EBITDA			0,61 x	1,06 x
Ratios Total pasivo / Patrimonio neto			2,69 x	4,24 x
Ratios de Cobertura al servicio de la deuda			0,72 x	0,99 x
Ratio corriente			1,35 x	1,21 x
			2016	
				0,76 x
				4,69 x
				0,99 x
				1,16 x

* Según cuentas individuales

Figura 12: Pantalla resumen de una empresa, con indicación del número de empleados

Proyecto: Herramienta de predicción de auditorías para empresas

Autores: ANR – JALF – JMR – MCP – MER – RNM

Versión: 2.0

Balance de situación		Cuenta de resultados		Individuales	
Principales Magnitudes del Balance y Cuenta de Resultados de PLATAFORMAS LOZANO SL - B41452855					
Activo (en €)	2018	2017	2016		
Activo no corriente	3.726.533	2.899.800	2.903.190		
Inmovilizado intangible	--	--	--		
Inmovilizado material	303.907	256.488	327.941		
Inversiones inmobiliarias	--	--	--		
Inversiones en empresas del grupo a largo plazo	3.353.964	2.574.910	2.508.865		
Inversiones financieras a largo plazo	68.662	68.202	66.384		
Otros activos a largo plazo	0	0	0		
Activo corriente	3.017.428	2.819.838	2.785.055		
Existencias	--	--	--		
Deudores comerciales	1.965.175	1.697.377	1.325.305		
Inversiones en empresas del grupo a corto plazo	653.963	896.827	943.267		
Inversiones financieras a corto plazo	--	--	68.993		
Otros activos a corto plazo	1.498	2.125	3.993		
Efectivo y activos líquidos equivalentes	396.791	223.510	443.497		
Total activo	6.743.960	5.719.438	5.688.245		
Pasivo (en €)	2018	2017	2016		
Patrimonio neto	2.509.204	1.349.790	1.214.089		
Fondos propios	2.509.204	1.349.790	1.214.089		
Capital	24.040	24.040	24.040		
Reservas	1.325.750	1.190.049	1.132.341		
Resultado del ejercicio	1.159.414	135.701	57.707		
Otros fondos propios	0	0	0		
Ajustes por cambio de valor	--	--	--		
Subvenciones y donaciones	--	--	--		
Pasivo no corriente	2.003.734	2.037.713	2.069.263		
Provisiones a largo plazo	--	--	--		
Deudas a largo plazo	3.734	37.713	69.263		
Deudas con empresas del grupo a largo plazo	2.000.000	2.000.000	2.000.000		
Otros pasivos a largo plazo	0	0	0		
Pasivo corriente	2.231.021	2.331.935	2.404.893		
Provisiones a corto plazo	--	--	--		
Deudas a corto plazo	472.309	252.403	134.078		
Deudas con empresas del grupo a corto plazo	670.413	1.113.750	1.620.212		
Acreedores comerciales	1.088.300	965.782	650.604		
Otros pasivos a corto plazo	0	0	0		
Total pasivo y patrimonio neto	6.743.960	5.719.438	5.688.245		

Figura 13: Balance de los 3 últimos años de la empresa, desglosado por partidas principales

Proyecto: Herramienta de predicción de auditorías para empresas

Autores: ANR – JALF – JMR – MCP – MER – RNM

Versión: 2.0

Balance de situación		Cuenta de resultados		
Principales Magnitudes del Balance y Cuenta de Resultados de PLATAFORMAS LOZANO SL - B41452855				
Cuenta de resultados (en €)	2018	2017	2016	
Ingresos de explotación	5.779.209	5.112.650	3.877.516	
Importe neto de la cifra de negocio	5.771.039	5.097.692	3.870.080	
Variación de existencias y productos en curso	--	--	--	
Trabajos realizados por el grupo para su activo	--	--	--	
Aprovisionamientos	-2.505.576	-2.450.540	-1.927.180	
Otros ingresos de explotación	8.170	14.958	7.436	
Gastos de personal	-1.719.781	-1.726.254	-1.176.543	
Otros gastos de explotación	-775.650	-702.374	-509.639	
Amortización del inmovilizado	-26.043	-30.095	-35.516	
Exceso de provisiones	--	--	--	
Otros resultados	4.423	38.941	3.790	
Resultado de explotación	756.583	242.328	232.428	
Ingresos financieros	25.420	32.601	12.715	
Gastos financieros	-110.198	-127.301	-135.285	
Diferencias de cambio	--	--	--	
Otros resultados financieros	779.054	66.045	20.415	
Resultado financiero	694.276	-28.655	-102.156	
Resultado antes de impuestos	1.450.859	213.673	130.272	
Impuestos sobre beneficios	-291.445	-77.972	-72.565	
Resultado del ejercicio	1.159.414	135.701	57.707	

Figura 14: Cuenta de explotación de los 3 últimos años de la empresa, desglosado por partidas principales

Por tanto, es una web candidata a ser *scrapeada*. Para ello se ha usado Octoparse, un software on-line. Al no haber las limitaciones de coste para adquirir los datos, se han obtenido los de empresas cuya facturación en el último año oscilaba entre 4.8 M euros y algo más de 10 M euros, para poder capturar también empresas que pudieran cumplir los criterios de Activo y Empleados, y no sólo el de Facturación. Por el límite superior de la facturación, entendemos que empresas que un año facturaron más de 10 M euros, es muy probable que el año anterior ya cumplieren con al menos 2 criterios de auditoría, por lo que se trataría de empresas que no solo caen fuera del objetivo de MLA², sino que además podrían contaminar el modelo al ser empresas de auditoría segura y lo que el modelo de MLA² quiere detectar es precisamente esas empresas que pasan de no precisar auditoría, a precisarla.

Con dichas empresas se ha generado un *dataset* con el modelo de datos de la Figura 8 más la extensión de la Tabla 8, con más de 10.000 empresas/registros.

Como desde un punto de vista de negocio lo que interesa es minimizar los falsos negativos, y dado que como se verá el modelo ganador es bastante preciso, se ha decidido poner el umbral en un 65% para la clase positiva, de tal manera que el AUC sea, como mínimo, superior al 85%. De esta manera, la probabilidad de que un cliente de MLA² acuda a una empresa que finalmente no precise auditoría es muy baja.

 Escuela de organización industrial	Programa Superior en Big Data & Business Analytics 7ªEd., Sevilla, Marzo – Septiembre 2020	
Proyecto: Herramienta de predicción de auditorías para empresas		
Autores: ANR – JALF – JMR – MCP – MER – RNM		Versión: 2.0

Con la ayuda de la herramienta on-line BigML, una solución de *Machine Learning As a Service*, se han aplicado varios modelos al *dataset* anterior, con los siguientes resultados:

- **Regresión logística:**

ACTUAL VS. PREDICTED	1	0	ACTUAL	RECALL	F	Phi
1	1,011	172	1,183	85.46%	0.86	0.65
0	169	678	847	80.05%	0.80	0.65
PREDICTED	1,180	850	2,030	82.75% AVG. RECALL	0.83 AVG. F	0.65 AVG. Phi
PRECISION	85.68%	79.76%	82.72% AVG. PRECISION	83.20% ACCURACY		

Figura 15: Matriz de confusión del modelo de regresión logística

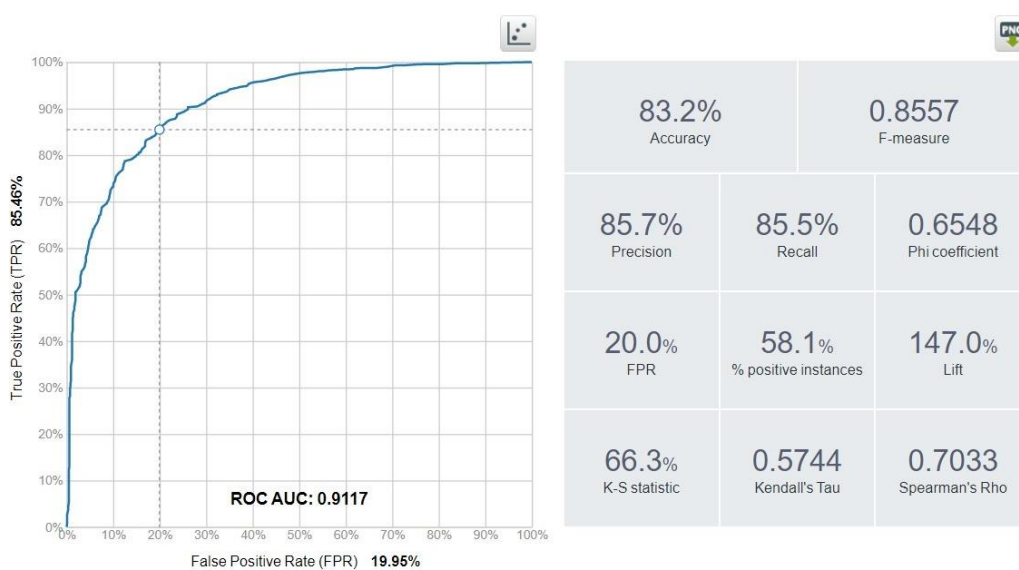


Figura 16: Métricas y curva ROC UAC del modelo de regresión logística

 Escuela de organización industrial	Programa Superior en Big Data & Business Analytics 7ªEd., Sevilla, Marzo – Septiembre 2020	
Proyecto: Herramienta de predicción de auditorías para empresas		
Autores: ANR – JALF – JMR – MCP – MER – RNM		Versión: 2.0

- **Random-forest:**

Un modelo de random-forest con 1.000 árboles proporciona los siguientes resultados:

ACTUAL VS. PREDICTED	1	0	ACTUAL	RECALL
1	1,139	44	1,183	96.28%
0	43	804	847	94.92%
PREDICTED	1,182	848	2,030	95.60% AVG. RECALL
PRECISION	96.36%	94.81%	95.59% AVG. PRECISION	95.71% ACCURACY

Figura 17: Matriz de confusión del modelo random-forest

95.7% Accuracy	0.9632 F-measure	
96.4% Precision	96.3% Recall	0.9119 Phi coefficient
5.1% FPR	58.2% % positive instances	165.4% Lift
91.8% K-S statistic	0.6855 Kendall's Tau	0.8357 Spearman's Rho

Figura 18: Métricas del modelo random-forest

Proyecto: Herramienta de predicción de auditorías para empresas

Autores: ANR – JALF – JMR – MCP – MER – RNM

Versión: 2.0

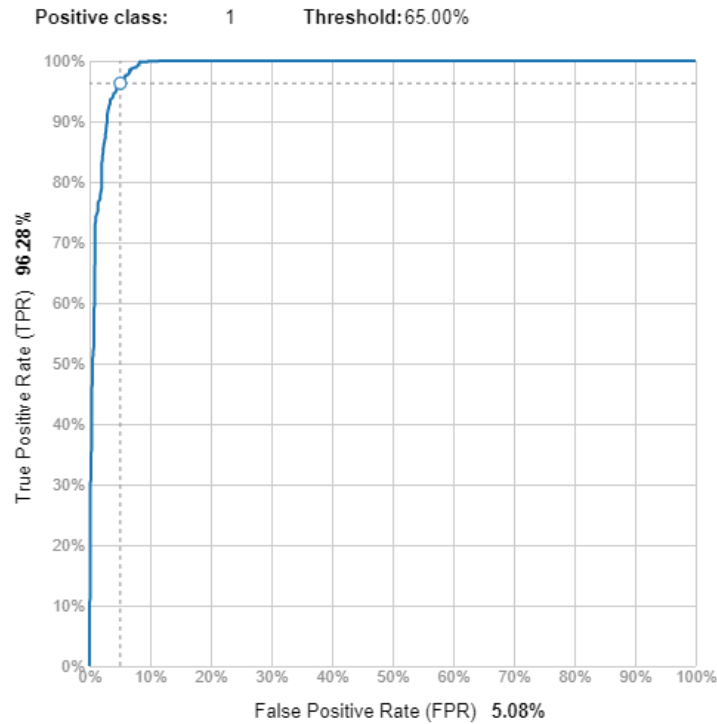


Figura 19: Curva ROC del modelo random forest con 1000 árboles

Como ya se anticipó en el apartado 5.3, con apenas 10 variables (un 5% del total considerado en el modelo), se puede explicar un 80% de la predicción, cumpliéndose una vez más la ley de Pareto¹¹:

¹¹ La Ley o Principio de Pareto, también conocida como la Regla del 80/20 (ó 20/80), es una ley empírica formulada por el ingeniero franco-italiano Wilfried Fritz Pareto que establece que, de forma general y para un amplio número de fenómenos, aproximadamente el 80% de las consecuencias proviene del 20% de las causas.

DatasetDepurado | Training (80%) Field Importances

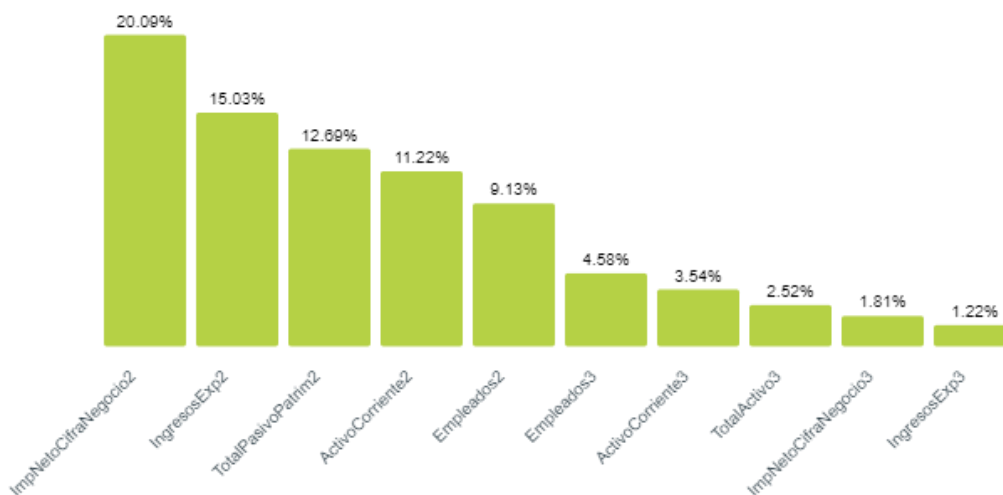


Figura 20: Peso de las 10 variables que justifican el 80% de la predicción

Estas variables son, por orden de importancia:

1. Importe neto de la cifra de negocio del último año cerrado: 20.09%
2. Ingresos de explotación del último año cerrado: 15.03%
3. Pasivo patrimonial total del último año cerrado: 12.69%
4. Activo corriente del último año cerrado: 11.22%
5. Empleados del último año cerrado: 9.13%
6. Empleados del penúltimo año cerrado: 4.58%
7. Activo del penúltimo año cerrado: 3.54%
8. Activo total del penúltimo año cerrado 3: 2.52%
9. Importe neto de la cifra de negocio del penúltimo año cerrado: 1.81%
10. Ingresos de explotación del penúltimo año cerrado: 1.22%

Proyecto: Herramienta de predicción de auditorías para empresas

Autores: ANR – JALF – JMR – MCP – MER – RNM

Versión: 2.0

• **Boosting:**

En teoría, este modelo debería dar mejores resultados. Y desde un punto de vista matemático, los da como se puede ver en las siguientes figuras:

ACTUAL VS. PREDICTED			ACTUAL	RECALL	F	Phi	
1	1,160	23	1,183	98.06%	0.97	0.92	
0	55	792	847	93.51%	0.95	0.92	
PREDICTED		1,215	815	2,030	95.78% AVG. RECALL	0.96 AVG. F	0.92 AVG. Phi
PRECISION		95.47%	97.18%	96.33% AVG. PRECISION	96.16% ACCURACY		

Figura 21: Matriz de confusión y métricas del modelo de ensembles con boosting



Figura 22: Curva ROC AUC y métricas del modelo de ensembles con boosting

No obstante, un análisis detallado de las variables que explican el resultado no permite dar mucha credibilidad a los resultados del modelo:

1. Otros ingresos de explotación del penúltimo año cerrado: 10.58%
2. Provincia: 9.16%
3. Importe neto de la cifra de negocio del último año cerrado: 5.29%
4. Activo total del último año cerrado: 3.75%
5. Ingresos de explotación del último año cerrado: 2.76%

	Programa Superior en Big Data & Business Analytics 7ªEd., Sevilla, Marzo – Septiembre 2020	
Proyecto: Herramienta de predicción de auditorías para empresas		
Autores: ANR – JALF – JMR – MCP – MER – RNM		Versión: 2.0

6. Ingresos financieros del último año cerrado: 2.45%
7. Empleados del último año cerrado: 2.39%
8. Nombre de la empresa: 2.22%
9. Otros ingresos de explotación del último año cerrado: 2.00%
10. Sector de actividad de la empresa: 1.96%

Es por ello que el modelo predictivo seleccionado ha sido el *random-forest*.

	Programa Superior en Big Data & Business Analytics 7ªEd., Sevilla, Marzo – Septiembre 2020	
Proyecto: Herramienta de predicción de auditorías para empresas		
Autores: ANR – JALF – JMR – MCP – MER – RNM		Versión: 2.0

6. Estudio Económico-Financiero

En el siguiente estudio económico-financiero se explican los balances de situación y cuentas de pérdidas y ganancias previstos para los tres primeros años de vida de MLA², y al mismo tiempo se presenta un análisis detallado de las inversiones y necesidades de financiación que tendrá el proyecto al inicio de su andadura.

Como se verá al final de dicho análisis, el proyecto no presenta problemas de caja y empieza a ser rentable a partir del año segundo de vida.

6.1 Principales clientes y previsión de ingresos

Teniendo en cuenta que el servicio ofrecido por MLA² tiene tres sabores diferentes y, además, incluye opciones con aspectos como exclusividad en los resultados, el cálculo de ingresos ha sido realizado generando un *mix* razonable de dichos servicios.

En concreto, se ha considerado que hasta la mitad de los clientes contratarían únicamente el paquete básico mientras que la otra mitad se repartiría entre los paquetes plus y premium a razón del 30% y del 20% respectivamente.

Por otro lado, el servicio de exclusividad ha sido concebido como algo básico para los clientes, dado que por la propia razón de ser de MLA² los clientes buscan precisamente ser los primeros en contactar con sus clientes para obtener / confirmar su ventaja competitiva por lo que ha sido considerado en el modelo de ingresos que hasta un 50% de nuestros clientes contratarán dicho servicio.

En cuanto a la evolución de clientes, también ha sido considerado un escenario bastante conservador en el que el primer mes de operación de la compañía se cerraría con tan solo 3 clientes y apenas 150 euros de ingresos, y a partir de ahí habría una progresión para acabar el año 1 con 319 clientes.

Estos datos han sido estimados teniendo en cuenta cual es el tamaño del mercado potencial de MLA² en España. A la fecha actual del segundo trimestre de 2020, y según información oficial procedente del ICAC (Instituto de Contabilidad y Auditoría de Cuentas), en España hay 2.253 auditores ejerciendo la profesión a título individual y 1.394 sociedades de auditoría, eso arroja un número máximo de clientes potenciales de 3.647. Esto significa que los 319 clientes considerados al cierre del ejercicio 1 son menos del 9% del mercado potencial, lo cual es muy razonable teniendo en cuenta,

	Programa Superior en Big Data & Business Analytics 7ª Ed., Sevilla, Marzo – Septiembre 2020	
Proyecto: Herramienta de predicción de auditorías para empresas		
Autores: ANR – JALF – JMR – MCP – MER – RNM		Versión: 2.0

como ha sido comentado anteriormente, que no existe un servicio similar en el mercado y que la competencia actual es inexistente.

	Programa Superior en Big Data & Business Analytics 7ªEd., Sevilla, Marzo – Septiembre 2020	
Proyecto: Herramienta de predicción de auditorías para empresas		
Autores: ANR – JALF – JMR – MCP – MER – RNM		Versión: 2.0

6.2 Plan de inversiones, gastos iniciales y balance final previsional

A continuación, se detallan las inversiones necesarias para la operación de MLA² durante los primeros años.

MLA² se constituirá con una aportación individual de 10.000 euros por cada uno de los seis socios que lo componen. Es decir, dispondrá de un capital social de inicio de 60.000 euros.

Aunque como se verá en el apartado 6.3 Plan de financiación, MLA² no presenta problemas de tesorería, se ha previsto la solicitud de un préstamo de 50.000 euros amortizable en 10 años para asegurar que no exista ninguna tensión de tesorería durante los primeros años de operación.

Se considera la compra de equipos informáticos para los cuatro empleados iniciales de MLA². Eso es los dos miembros del equipo técnico, más el director comercial y el director ejecutivo.

Proyecto: Herramienta de predicción de auditorías para empresas

Autores: ANR – JALF – JMR – MCP – MER – RNM

Versión: 2.0

ACTIVO	2021	2022	2023
A) ACTIVO NO CORRIENTE	2.250,00 €	1.500,00 €	750,00 €
I. Inmovilizado intangible	0,00 €	0,00 €	0,00 €
II. Bienes del Patrimonio Histórico	0,00 €	0,00 €	0,00 €
III. Inmovilizado material	2.250,00 €	1.500,00 €	750,00 €
IV. Inversiones inmobiliarias	0,00 €	0,00 €	0,00 €
V. Inversiones en entidades del grupo y asociadas a largo plazo	0,00 €	0,00 €	0,00 €
VI. Inversiones financieras a largo plazo	0,00 €	0,00 €	0,00 €
VII. Activos por impuesto diferido (1)	0,00 €	0,00 €	0,00 €
B) ACTIVO CORRIENTE	120.185,23 €	181.207,83 €	298.158,63 €
I. Existencias	0,00 €	0,00 €	0,00 €
II. Usuarios y otros deudores de la actividad propia	0,00 €	0,00 €	0,00 €
III. Deudores comerciales y otras cuentas a cobrar	16.050,12 €	20.865,15 €	27.124,70 €
IV. Inversiones en entidades del grupo y asociadas a corto plazo	0,00 €	0,00 €	0,00 €
V. Inversiones financieras a corto plazo	0,00 €	0,00 €	0,00 €
VI. Periodificaciones a corto plazo	0,00 €	0,00 €	0,00 €
VII. Efectivo y otros activos líquidos equivalentes	104.135,11 €	160.342,68 €	271.033,93 €
TOTAL ACTIVO (A+B)	122.435,23 €	182.707,83 €	298.908,63 €
PASIVO	2021	2022	2023
A) PATRIMONIO NETO	60.506,23 €	109.507,24 €	212.085,65 €
A-1) Fondos propios	60.506,23 €	109.507,24 €	212.085,65 €
I. Capital social	60.000,00 €	60.000,00 €	60.000,00 €
1. Capital social	60.000,00 €	60.000,00 €	60.000,00 €
2.(C.S. no exigido)	0,00 €	0,00 €	0,00 €
II. Reservas	0,00 €	506,23 €	49.507,24 €
III. Excedentes de ejercicios anteriores	0,00 €	0,00 €	0,00 €
IV. Excedente del ejercicio	506,23 €	49.001,01 €	102.578,41 €
A-2) Subvenciones, donaciones y legados recibidos	0,00 €	0,00 €	0,00 €
B) PASIVO NO CORRIENTE	41.530,42 €	37.035,11 €	32.356,65 €
I. Provisiones a largo plazo	0,00 €	0,00 €	0,00 €
II. Deudas a largo plazo	41.530,42 €	37.035,11 €	32.356,65 €
1. Deudas con entidades de crédito	41.530,42 €	37.035,11 €	32.356,65 €
2. Acreedores por arrendamiento financiero	0,00 €	0,00 €	0,00 €
3. Otras deudas a largo plazo	0,00 €	0,00 €	0,00 €
III. Deudas con entidades del grupo y asociadas a largo plazo	0,00 €	0,00 €	0,00 €
IV. Pasivos por impuesto diferido (1)	0,00 €	0,00 €	0,00 €
V. Periodificaciones a largo plazo	0,00 €	0,00 €	0,00 €
C) PASIVO CORRIENTE	20.398,57 €	36.165,48 €	54.466,32 €
I. Provisiones a corto plazo	0,00 €	0,00 €	0,00 €
II. Deudas a corto plazo	16.994,34 €	17.423,81 €	17.865,53 €
1. Deudas con entidades de crédito	4.319,34 €	4.495,31 €	4.678,46 €
2. Acreedores por arrendamiento financiero	0,00 €	0,00 €	0,00 €
3. Otras deudas a corto plazo	12.675,00 €	12.928,50 €	13.187,07 €
III. Deudas con entidades del grupo y asociadas a corto plazo	0,00 €	0,00 €	0,00 €
V. Acreedores comerciales y otras cuentas a pagar	3.404,24 €	18.741,66 €	36.600,80 €
1. Proveedores	3.010,49 €	2.182,99 €	2.182,99 €
2. Otros acreedores	393,74 €	16.558,67 €	34.417,80 €
VI. Periodificaciones a corto plazo	0,00 €	0,00 €	0,00 €
TOTAL PATRIMONIO NETO Y PASIVO (A+B+C)	122.435,23 €	182.707,83 €	298.908,63 €

Figura 23: Balance previsional del MLA² a 3 años

	Programa Superior en Big Data & Business Analytics 7ªEd., Sevilla, Marzo – Septiembre 2020	
Proyecto: Herramienta de predicción de auditorías para empresas		
Autores: ANR – JALF – JMR – MCP – MER – RNM		Versión: 2.0

6.3 Plan de financiación

El objetivo principal de la estrategia financiera pasa por obtener un fondo de maniobra elevado, con el que poder hacer frente a cualquier contratiempo de tesorería que pudiera tener MLA². Asimismo, la buena capacidad de liquidez a corto plazo proporciona un buen colchón económico con el que soportar la solvencia interna de la empresa en el posible caso de una evolución (no tan certera como se ha pronosticado) de la captación de clientes.

La estrategia financiera de MLA² pasa por recurrir a un préstamo bancario que inyecte liquidez en el primer ejercicio de la actividad. Si bien existen opciones que pueden ser más beneficiosas para obtener un préstamo con entidades dedicadas a la aceleración de proyectos tecnológicos, se ha considerado la opción de una entidad financiera corriente, para simplificar el modelo de negocio inicial.

Este préstamo tendrá un capital de 50.000 euros, un plazo de duración de 10 años, pudiendo ser amortizado en cualquier momento sin un coste elevado, y se espera conseguir a un tipo de interés aproximado a la media del mercado actual, que podría situarse en el 4% TIN / 4.252% TAE. En el cálculo del citado préstamo se ha determinado una cuota de pago mensual de 506 euros lineal y constante.

MLA² tiene como objetivo de ingresos obtener un tercio de la cuota de mercado de posibles clientes, en los primeros 3-4 años desde su nacimiento/lanzamiento. El objetivo financiero es cancelar el préstamo con la entidad bancaria a partir del 4º año, una vez alcanzado el objetivo de ingresos.

De esta forma, la rentabilidad económica y financiera sería positiva sin financiación ajena a partir del 4º año, lo que supone un plan económico atractivo para un mercado (captación de clientes de auditoría) que está por abrir y que tiene una gran escalabilidad de negocio.

La captación de otras fuentes de financiación como subvenciones de ayudas a *startup* o PYMES, así como contratos de publicidad, promoción y/o patrocinadores se contemplan durante el primer ejercicio de MLA². Un ejemplo de subvenciones puede ser el programa SME Instrument de la UE, que ofrece subvenciones para PYMES innovadoras en las distintas fases de desarrollo de su proyecto.

En un pequeño sondeo de mercado podemos pensar que instituciones como el Instituto de Contabilidad y Auditores de Cuentas (ICAC), el Instituto de Censores

	Programa Superior en Big Data & Business Analytics 7ªEd., Sevilla, Marzo – Septiembre 2020	
Proyecto: Herramienta de predicción de auditorías para empresas		
Autores: ANR – JALF – JMR – MCP – MER – RNM		Versión: 2.0

Jurados de Cuentas de España (ICJCE), el Colegio de Economistas o proveedores de datos como Einforma, InfoCif, etc, puedan estar interesados en realizar contratos de patrocinio y promoción en nuestra web. Estas operaciones, además de dar un impulso financiero a la cuenta de pérdidas y ganancias de MLA² en el primer ejercicio, pueden suponer una buena estrategia de marketing con la que obtener fiabilidad en el mercado de cara a nuestros clientes, los auditores, quienes conocen perfectamente a tales instituciones y empresas, que gozan de una fuerte credibilidad.

En búsqueda de financiación externa también existen aceleradoras de proyectos o *seed capital*, si bien en España es necesario contar con un plan de negocios bien definido o incluso que la empresa haya dado ya sus primeros pasos y cuente con algunos clientes, por lo que se considera que esta opción puede buscarse en un futuro si fuese necesaria. CDTI, DAD, Caixa Capital Risc y BStartup, AIEI, Seed Rocket son algunas de ellas.

	Programa Superior en Big Data & Business Analytics 7ªEd., Sevilla, Marzo – Septiembre 2020	
Proyecto: Herramienta de predicción de auditorías para empresas		
Autores: ANR – JALF – JMR – MCP – MER – RNM		Versión: 2.0

6.4 Cuentas de pérdidas y ganancias

Se presenta a continuación el desglose de la cuenta de pérdidas y ganancias:

- **Ingresos**

- Ingresos por cuotas de clientes
- Ingresos de promociones, patrocinadores y colaboraciones: se ha estimado que MLA² será capaz de generar 15.000 euros al año por publicidad en nuestra herramienta de uso del auditor. Esta publicidad podría venir por parte de eInforma, al estar los datos extraídos de su sistema, o por parte de los proveedores recurrentes de las empresas de auditoría a los que se contactará para ofrecerle estos espacios publicitarios.
- Otros ingresos de la actividad: ha sido considerado un importe 3.000 euros durante el año 1 en concepto de subvenciones a conseguir. Ese importe será invertido concretamente en la compra del material informático para los empleados. Se ve reflejada por tanto con signo negativo también tanto en la partida “otros gastos” como en la partida de “amortización del inmovilizado”.

- **Gastos**

- Aprovisionamientos: esta partida se compone principalmente de los gastos derivados del uso de la plataforma tecnológica Azure en la que se basa el servicio, y que ha sido explicada en detalle a lo largo del punto 5, el gasto en datos procedentes de eInforma y una pequeña partida en gastos de ofimática, telefonía y mantenimiento del sitio web de MLA².
- Gastos financieros: estos gastos son los derivados de los intereses correspondientes al préstamo de 50.000 euros comentado anteriormente.
- Otros gastos de la actividad: en este apartado se concentran tres partidas:
 - Gastos de representación: estos son los gastos estimados para que el director comercial desarrolle su actividad de desarrollo de negocio, ya sea para gastos de viaje, pago de eventos, etc.

Proyecto: Herramienta de predicción de auditorías para empresas

Autores: ANR – JALF – JMR – MCP – MER – RNM

Versión: 2.0

- Gestoría y seguros de crédito y caución: si bien, dados los perfiles de los socios que componen MLA², la mayor parte de las tareas administrativas como contabilidad, liquidación de impuestos, etc. será realizada por ellos, se ha considerado esta para incluir los gastos potenciales asociados a una gestoría que ayude a los socios. Adicionalmente, en este bloque se ha considerado también la contratación de un seguro de crédito y caución, estimado en el 1% del valor de las ventas para asegurar el cobro de las partidas de clientes.

CUENTA DE RESULTADOS	(Debe)		Haber
	2021	2022	2023
A) Excedente del ejercicio			
1. Ingresos de la actividad propia	15.000,00	15.000,00	15.000,00
a) Cuotas de asociados y afiliados	0,00	0,00	0,00
b) Aportaciones de usuarios	0,00	0,00	0,00
c) Ingresos de promociones, patrocinadores y colaboraciones	15.000,00	15.000,00	15.000,00
d) Subvenciones, donaciones y legados imputados al excedente del ejercicio	0,00	0,00	0,00
e) Reintegro de ayudas y asignaciones	0,00	0,00	0,00
2. INCN	192.601,39	250.381,80	325.496,34
a) Estandar	83.648,40	108.742,91	141.365,79
b) Plus	50.189,04	65.245,75	84.819,47
c) Premium	33.459,36	43.497,17	56.546,32
d) Exclusividad	25.304,59	32.895,97	42.764,76
3. Gastos por ayudas y otros	0,00	0,00	0,00
5. Trabajos realizados por la entidad para su activo	0,00	0,00	0,00
6. Aprovisionamientos	-38.825,93	-28.895,93	-28.895,93
a) Einforma	-19.860,00	-9.930,00	-9.930,00
b) Azure	-16.265,93	-16.265,93	-16.265,93
c) Gastos ofimática, web y telefonía	-2.700,00	-2.700,00	-2.700,00
7. Otros ingresos de la actividad	3.000,00	0,00	0,00
8. Gastos de personal	-152.100,00	-155.142,00	-158.244,84
9. Otros gastos de la actividad	-15.926,01	-13.503,82	-14.254,96
a) Gastos representación	-10.000,00	-10.000,00	-10.000,00
b) Otros gastos	-3.000,00	0,00	0,00
c) Gestoría y seguros crédito y caución	-2.926,01	-3.503,82	-4.254,96
10. Amortización del inmovilizado	-750,00	-750,00	-750,00
11. Subvenciones, donaciones y legados de capital traspasados al excedente del ejercicio	0,00	0,00	0,00
12. Excesos de provisiones	0,00	0,00	0,00
13. Deterioro y resultado por enajenación de inmovilizado	0,00	0,00	0,00
A.1) EXCEDENTE DE LA ACTIVIDAD (1+2+3+4+5+6+7+8+9+10+11+12+13)	2.999,44	67.090,05	138.350,61
14. Ingresos financieros	0,00	0,00	0,00
15. Gastos financieros	-2.324,46	-1.755,37	-1.579,40
16. Variación de valor razonable en instrumentos financieros	0,00	0,00	0,00
17. Diferencias de cambio	0,00	0,00	0,00
18. Deterioro y resultado por enajenaciones de instrumentos financieros	0,00	0,00	0,00
A.2) EXCEDENTE DE LAS OPERACIONES FINANCIERAS (14+15+16+17+18)	-2.324,46	-1.755,37	-1.579,40
A.3) EXCEDENTE ANTES DE IMPUESTOS (A.1+A.2)	674,98	65.334,68	136.771,21
19. Impuestos sobre beneficios	-168,74	-16.333,67	-34.192,80
A.4) Variación de patrimonio neto reconocida en el excedente del ejercicio (A.3+19)	506,23	49.001,01	102.578,41
I) RESULTADO TOTAL, VARIACIÓN DEL PATRIMONIO NETO EN EL EJERCICIO (A.4+D+E+F+G+H)	506,23	49.001,01	102.578,41

Figura 24: Cuenta de previsional del pérdidas y ganancias de MLA² a 3 años

6.5 Plan de Tesorería

La Figura 25 muestra las necesidades de financiación de MLA², sin tener en cuenta el préstamo financiero y capital invertido inicial.

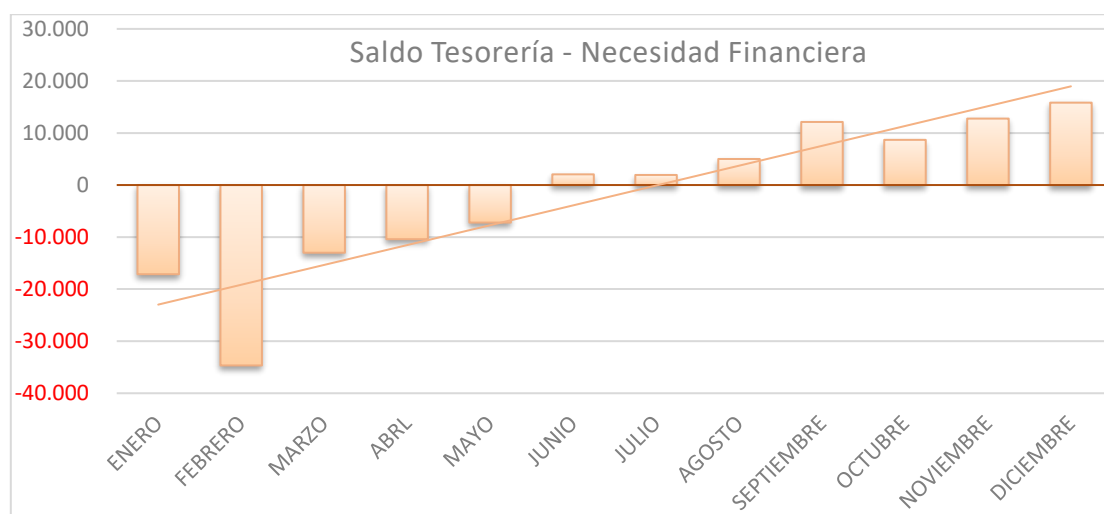


Figura 25:Saldo de tesorería y (necesidad financiera)

A continuación, la Figura 26 muestra la evolución del saldo de caja durante los primeros 12 meses, incluyendo préstamo bancario y el desembolso del capital invertido inicial por los socios.

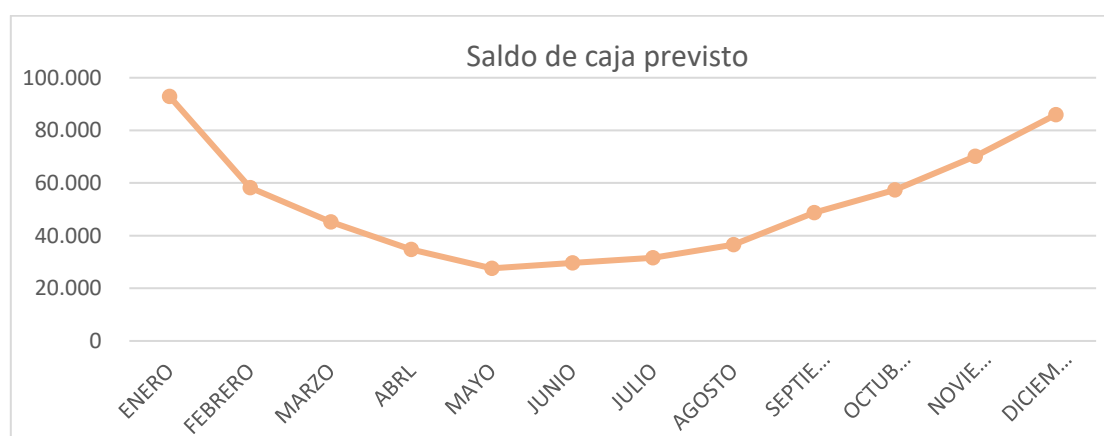


Figura 26:Previsión de saldo de caja

Por último, en la Figura 27 se puede ver el detalle del Plan de tesorería de MLA², en el que se puede constatar la viabilidad del proyecto y cómo las tensiones de tesorería disminuyen a partir del sexto mes de funcionamiento del proyecto.

Proyecto: Herramienta de predicción de auditorías para empresas

Autores: ANR – JALF – JMR – MCP – MER – RNM

Versión: 2.0

MES	ENERO	FEBRERO	MARZO	ABRIL	MAYO	JUNIO	JULIO	AGOSTO	SEPTIEMBRE	OCTUBRE	NOVIEMBRE	DICIEMBRE
1	SALDO INICIAL 110.000,00											
2	COBROS											
<i>Promociones / Patrocinadores</i>						5.000,00			5.000,00			5.000,00
<i>Ingresos por cuotas</i>	0,00	219,00	2.737,50	4.562,50	7.300,00	10.950,00	15.141,89	17.793,75	19.618,75	20.965,69	21.900,00	22.812,50
<i>Ingresos por exclusividad</i>	0,00	0,00	0,00	801,74	1.282,79	1.924,18	2.660,80	3.126,80	3.447,49	3.684,18	3.848,36	4.008,71
<i>Otros ingresos de la actividad</i>											3.000,00	
TOTAL	0,00	219,00	2.737,50	5.364,24	8.582,79	17.874,18	17.802,69	20.920,55	28.066,24	24.649,88	28.748,36	31.821,21
3	PAGOS											
<i>elinforma</i>	0,00	19.200,00	60,00	60,00	60,00	60,00	60,00	60,00	60,00	60,00	60,00	60,00
<i>Azure</i>	0,00	1.357,96	1.355,27	1.355,27	1.355,27	1.355,27	1.355,27	1.355,27	1.355,27	1.355,27	1.355,27	1.355,27
<i>Gastos ofimática, web y telefonía</i>	225,00	225,00	225,00	225,00	225,00	225,00	225,00	225,00	225,00	225,00	225,00	225,00
<i>Gastos de personal</i>	12.675,00	12.675,00	12.675,00	12.675,00	12.675,00	12.675,00	12.675,00	12.675,00	12.675,00	12.675,00	12.675,00	12.675,00
<i>Gastos de representación</i>	833,33	833,33	833,33	833,33	833,33	833,33	833,33	833,33	833,33	833,33	833,33	833,33
<i>Otros gastos</i>	3.000,00											
<i>Gestoría y seguros crédito y caución</i>	0,00	83,33	85,52	110,71	136,98	169,16	212,08	261,36	292,54	314,00	329,83	340,82
<i>Amortización préstamo</i>	400,00	506,23	506,23	506,23	506,23	506,23	506,23	506,23	506,23	506,23	506,23	506,23
TOTAL	17.133,33	34.880,85	15.740,35	15.765,54	15.791,80	15.823,99	15.866,90	15.916,19	15.947,37	15.968,82	15.984,66	15.995,65
4	SALDO mes											
5	SALDO acumulado											
6	SALDO CAJA											

Figura 27: Plan de tesorería del primer año

	Programa Superior en Big Data & Business Analytics 7ªEd., Sevilla, Marzo – Septiembre 2020	
Proyecto: Herramienta de predicción de auditorías para empresas		
Autores: ANR – JALF – JMR – MCP – MER – RNM		Versión: 2.0

6.6 Ratios y principales indicadores financieros

Con los ratios económicos podremos comprobar la rentabilidad del negocio desde distintos puntos de vista y se parte, fundamentalmente, de los datos contenidos en la cuenta de pérdidas y ganancias. Con un análisis económico valoraremos aspectos como el margen bruto que se obtiene sobre las ventas, la rentabilidad de los activos o la rotación del activo circulante o a corto plazo.

Con los ratios financieros estudiaremos los recursos propios y ajenos con los que se pretende financiar el proyecto, éstos se calculan sobre los datos del balance de situación. Mediante el análisis financiero podremos determinar por tanto el nivel de endeudamiento de la entidad o su grado de solvencia.

Estos son los principales ratios económicos y financieros:

RESULTADOS	2021	2022	2023
Ratio de Endeudamiento	0,3	0,1	0,1
R. Solvencia	2,0	2,5	3,4
R. Tesorería	5,9	5,0	5,5
Grado Apalancamiento Financ.	0,5	1,6	1,4
Fondo de Maniobra	99.786,7 €	145.042,4 €	243.692,3 €
Margen Bruto sobre Ventas	1,56%	26,80%	42,50%
Rotación Activo Circulante	1,6	1,4	1,1
Rentabilidad Económica	2%	37%	46%
Rentabilidad Financiera	1%	45%	48%

Tabla 9: Previsiones a 3 años de los principales ratios económicos y financieros de MLA²

De los resultados obtenidos, podemos obtener las siguientes conclusiones:

- El Ratio de Endeudamiento determina que el proyecto de MLA² tal como está concebido, no está soportando un volumen elevado de deuda y por consiguiente no existe descapitalización. Sí puede estar incurriéndose en un exceso de capitales ociosos, que puede afectar a la rentabilidad de los recursos. Esto se debe principalmente a la estrategia utilizada para tener efectivo en caja

	Programa Superior en Big Data & Business Analytics 7ªEd., Sevilla, Marzo – Septiembre 2020	
Proyecto: Herramienta de predicción de auditorías para empresas		
Autores: ANR – JALF – JMR – MCP – MER – RNM		Versión: 2.0

a corto plazo en los primeros años de vida del proyecto, mediante la solicitud del préstamo bancario.

- El Ratio de Solvencia mide la capacidad de MLA² para hacer frente a las deudas. Como se puede comprobar, es superior a la unidad, lo que indica fuerte capacidad de solvencia.
- El Ratio de Tesorería compara el efectivo en caja y bancos, y los bienes y derechos que se pueden convertir en dinero con mucha rapidez (principalmente clientes y deudores a corto plazo) con las deudas a responder en el corto plazo. Este ratio es muy positivo.
- El Grado de Apalancamiento Financiero indica la rentabilidad de acudir a la financiación ajena. En el primer ejercicio, el coste de la financiación ajena es superior a la financiación propia y esto genera una menor rentabilidad. En los ejercicios siguientes, la financiación ajena tiene una rentabilidad elevada y por tanto corrobora el acierto de la estrategia de solicitud del préstamo.
- El Fondo de Maniobra pone de manifiesto la naturaleza de los recursos que financian el activo corriente para determinar, con su análisis, posibles tensiones de tesorería que puede tener la empresa, problemas más o menos graves de liquidez o, por el contrario, su buena salud financiera. El principal objetivo de la solicitud del préstamo es el resultado tan positivo del citado indicador.
- El Margen Bruto sobre las Ventas tiene un crecimiento exponencial debido al incremento del resultado de explotación. Indica el % de beneficio extraído de los ingresos netos que se obtienen en la actividad del negocio.
- La Rotación del Activo Circulante refleja las veces que se ha utilizado el mismo, en este caso efectivo y el disponible/realizable, en la obtención de las ventas. Su valor superior a la unidad indica un buen aprovechamiento de los recursos disponibles.
- La Rentabilidad Económica indica que se requieren pocos activos para lograr un beneficio alto, principalmente a partir del ejercicio 2.
- La Rentabilidad Financiera mide la remuneración a los propietarios de la empresa. Cuanto más alto sea, mayores serán los beneficios obtenidos por los inversores. Indica, al igual que la rentabilidad económica, que a partir del ejercicio 2 el proyecto remunera positivamente a los inversores.

	Programa Superior en Big Data & Business Analytics 7ªEd., Sevilla, Marzo – Septiembre 2020	
Proyecto: Herramienta de predicción de auditorías para empresas		
Autores: ANR – JALF – JMR – MCP – MER – RNM		Versión: 2.0

7. Conclusiones

Como se ha podido ver a lo largo del documento, se presenta una idea de empresa que, partiendo de una realidad existente en el sector de la auditoría, como es el hecho de que se trata de un mercado muy maduro, con alta competencia y con unos márgenes muy ajustados, trata de democratizar el sector haciendo para ello uso de dos piezas claves:

- Sus socios: un equipo altamente preparado y con un conocimiento muy amplio del sector de la auditoría y de la tecnología.
- El Big Data: la introducción de esta tecnología en este mercado será clave para la consecución de este objetivo.

El proyecto presenta una estructura económica muy sólida y con una generación de caja que garantiza la rentabilidad futura del mismo.

Por último, hay que recordar que no existe actualmente competencia en el mercado que proporcione un servicio similar.