SonaR

OPTIMIZACIÓN DEL PROCESO DE ASIGNACIÓN DE RECURSOS PARA LA RESOLUCIÓN DE INCIDENCIAS

Índice

1. Introducción

Cuando se habla de prestación de servicios, conseguir el balance óptimo entre los recursos empleados para ofrecer un buen nivel de servicio y la eficiencia económica y de procesos, suele ser una tarea complicada en la mayoría de los sectores económicos, incluidos el sector de las telecomunicaciones y de las tecnologías de la información y las comunicaciones. De este problema no escapa INDEA, una empresa dirigida a la prestación de servicios de instalación y mantenimiento de equipos y servicios de cliente en el sector de telecomunicaciones, electrónica y tecnologías de la información, especializada en la instalación de redes y equipos de telecomunicaciones.

En particular, en el mundo de las telecomunicaciones y de las tecnologías de la información, la consecución de este balance implica un alto esfuerzo humano y tecnológico, que se encuentra muchas veces atado a personas y conocimientos específicos los cuales, en caso de no estar presentes, afectan de forma significativa a la calidad y efectividad de los servicios prestados. El presente proyecto se enmarca en un marco de colaboración entre la Dirección de Sistemas de Información de INDEA y SonaR, equipo conformado por Andrés Ortiz, David Crespo, Isabel Martins y Jesús Ruiz.

1.1. Contexto

En la actualidad, Indea gestiona servicios de instalación y mantenimiento de equipos de telecomunicaciones y electrónica para diversas empresas, entre las que destacan, por ejemplo, el Grupo Ingénico, Vodafone, ONO y Nokia. El presente proyecto, tal y como se argumentará más adelante, se centrará en la **optimización del proceso de asignación de recursos de Indea para la resolución de incidencias del Grupo Ingénico**.

Indea Ingeniería de Aplicaciones Sociedad Limitada (en adelante, Indea) es una empresa ubicada en Paterna, Valencia, con ámbito geográfico de actuación nacional. Constituida a principios de 2004, es una empresa dedicada a prestar servicios de telecomunicaciones, encontrándose entre las que más factura en su sector, tal y como acredita el sello TOP 100.000 empresas. En concreto, Indea presta servicios de instalación y mantenimiento de equipos y servicios de cliente en el sector de telecomunicaciones, electrónica y tecnologías de la información. Ofrece un conjunto integral de servicios que abarcan la ingeniería y desarrollo de proyectos, el despliegue de red, la instalación de equipos en el cliente final y mantenimiento. Dentro de su división de I+M también ha desarrollado otras líneas actuación complementarias, como Sistemas de seguridad y videovigilancia sobre IP, TPVs o cajeros automáticos. Disponen de delegaciones y almacenes en varias provincias para poder dar servicio a los clientes en todo el territorio nacional. Algunos de sus principales clientes en la actualidad

son Ingenico, Vodafone, ONO o Nokia. En relación a su información financiera, a continuación se muestran dos imágenes con datos financieros extraídos de SABI (EOI); en el anexo III se incluye el informe completo.

Información legal & tipo cuentas

Forma jurídica	Sociedad limitada	**Ultimo año disponible**	31/12/2016
Forma jurídica detallada	Sociedad limitada	**Años disponibles**	13
Capital social (EUR)	3.006	**Cuentas disponibles**	No Consolidadas
Fecha constitución	09/01/2004		
Estado	Activa		
Estado detallado	Activa		
Director ejecutivo	Don Rafael Martinez Luna		

Información grupo & tamaño (2016)

Ingresos explotación	7.927.950 EUR	**Indicador Independencia BvD**	D
Resultado ejercicio	12.940 EUR	**Empresas en el grupo corporativo**	2
Total activo	4.921.798 EUR	**Núm. accionistas**	1
Número de empleados	100	**Núm. participadas**	1

Ilustración 1. Información legal de Indea (fuente: SABI, accedido 30 Marzo 2019)

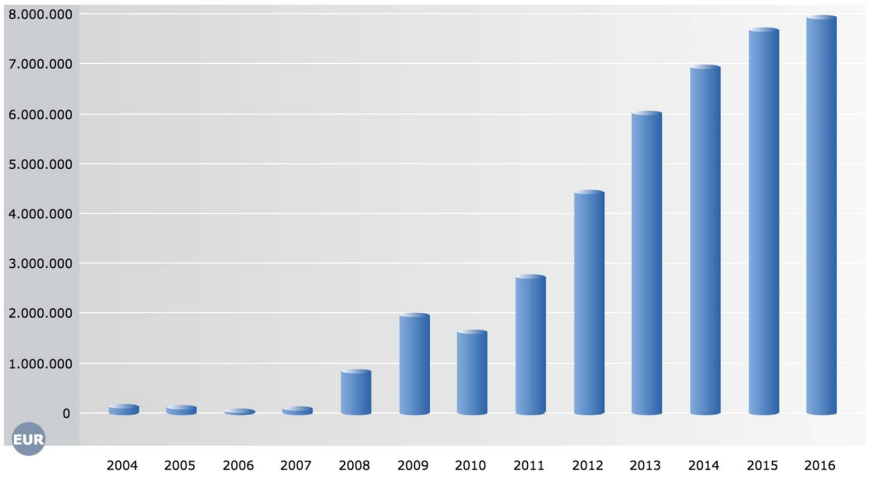


Ilustración 2. Evolución de los ingresos por explotación de Indea durante los últimos años.

Por su parte, el Grupo Ingénico es una empresa francesa, líder en el sector de pagos, enfocada en proveer a sus clientes de medios seguros e inteligentes para que los comerciantes puedan gestionar los pagos de sus consumidores. En España, son el principal proveedor de Terminales de Punto de Venta (TPV) de los bancos, y, por lo tanto, sus terminales están instalados en la mayoría de los comercios españoles que cuentan con servicio de punto de venta.

Cuando un comercio solicita a un banco la instalación de un TPV, el banco en cuestión realiza una petición al Grupo Ingénico, quienes externalizan la instalación de los terminales a INDEA. Para realizar la asignación de las diferentes solicitudes de instalaciones y mantenimientos, el Grupo Ingénico cuenta

con una plataforma de *tickets*, a través de la cual controla las incidencias pendientes de atención, entendiéndose por incidencia toda aquella nueva solicitud de instalación así como todas aquellas solicitudes de mantenimiento de TPVs actualmente instalados en el comercio final.

Para la gestión de dichos servicios de instalación y mantenimiento, INDEA cuenta con un software interno creado por su departamento informático (denominado “Pez”), mediante el cual se conecta a la plataforma del Grupo Ingénico y recolecta los tickets existentes pendientes de atención. En la figura a continuación, se muestra el ciclo de vida de un ticket en el sistema INDEA.

Ilustración 3: Ciclo de vida de un ticket en el sistema Indea

Una vez que los tickets han sido cargados en el sistema interno de INDEA, un grupo de coordinadores procede a realizar la asignación manual de las incidencias a los diferentes técnicos de campo, teniendo en cuenta factores como: la geolocalización del técnico, si el técnico tiene o no el equipo que tiene que ser instalado, la experiencia del técnico, etc. Esta asignación ha de realizarse de la mejor manera posible para poder cumplir el SLA (*Service Level Agreement*) acordado con el Grupo Ingénico, que actualmente es del 98%, considerándose un KPI de 24 horas para la resolución de tickets asociados a mantenimientos, y un KPI de 48 horas para la resolución de tickets asociados a nuevas instalaciones (ver Ilustración 2).

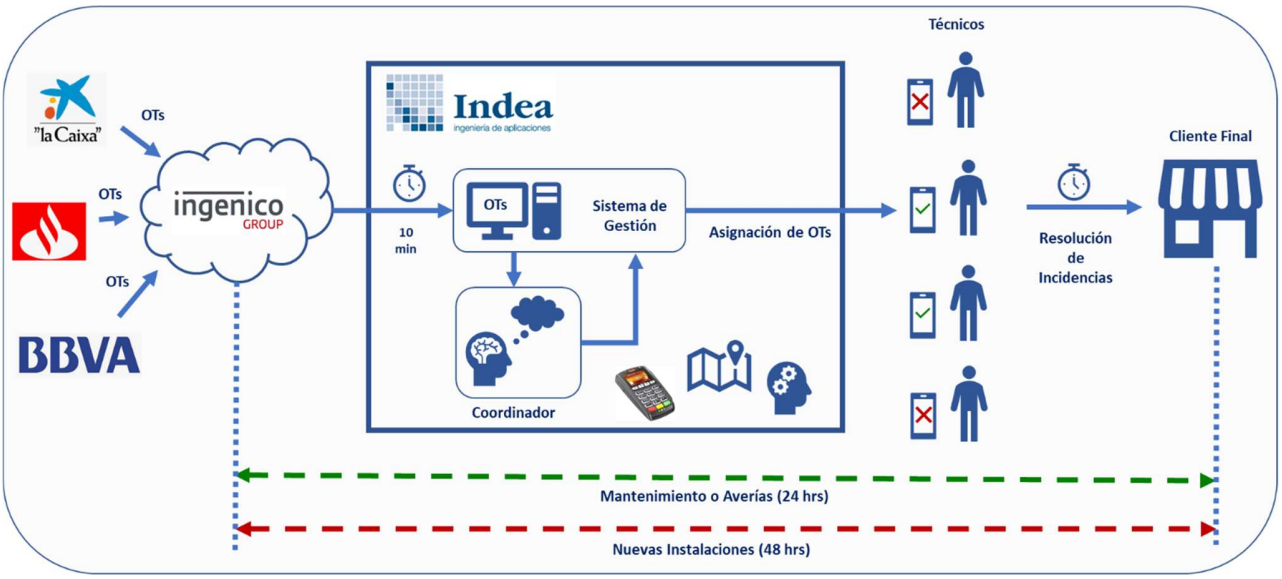


Ilustración 4. Contexto del problema

El proceso de asignación de técnicos a resolución de incidencias descrito anteriormente muestra una dependencia muy alta de la gestión manual de los coordinadores en el cumplimiento de los SLA, representando un problema para INDEA cuando uno de estos coordinadores se va de vacaciones, baja o simplemente renuncia a su trabajo.

Adicionalmente, aún cuando INDEA tiene un nivel de cumplimiento del SLA de un 97% (lo cual supone un incumplimiento del 1% respecto al SLA contratado, mencionado anteriormente), conseguirlo le cuesta mucho esfuerzo y recursos, y la expectativa actual de INDEA es manejar el creciente número de incidencias sin tener que crecer en número de recursos.

1.2. Identificación y priorización de causas del problema

A la hora de identificar y priorizar las causas a intervenir es importante basarse en hechos reales y objetivos. Asimismo, surge la necesidad de aplicar técnicas de solución de problemas adecuadas. Para el caso de INDEA se han aplicado las siguientes técnicas en el análisis del problema: entrevistas, recolección de datos, *brainstorming* y diagrama de *Ishikawa*.

- Entrevistas

Esta técnica ha permitido recopilar información directamente de las personas involucrados en los procesos implicados en el problema. Para el proyecto, se ha contactado con D. Diego Piedrahita, responsable de despliegue de todas las regiones en Indea y Director de Sistemas de Información de la compañía, formando parte de su Consejo de Dirección. Esta persona no sólo ha sido el contacto con la empresa, sino que ha actuado en todo momento como facilitador y promotor del proyecto en el cliente.

Si bien han realizado numerosas comunicaciones con esta persona y miembros de su equipo, a través de diversos canales (mail, teléfono, whatsApp, etc.), se han realizado 2 entrevistas formales en las que se ha analizado con mayor profundidad el problema y los planteamientos posibles. Las notas de dichas entrevistas se adjuntan en el Anexo I de la presente memoria.

- Diagrama de Ishikawa

  Se ha utilizado esta técnica de análisis para encontrar las causas origen del problema. Se presenta a continuación el diagrama elaborado:

- Recolección de datos

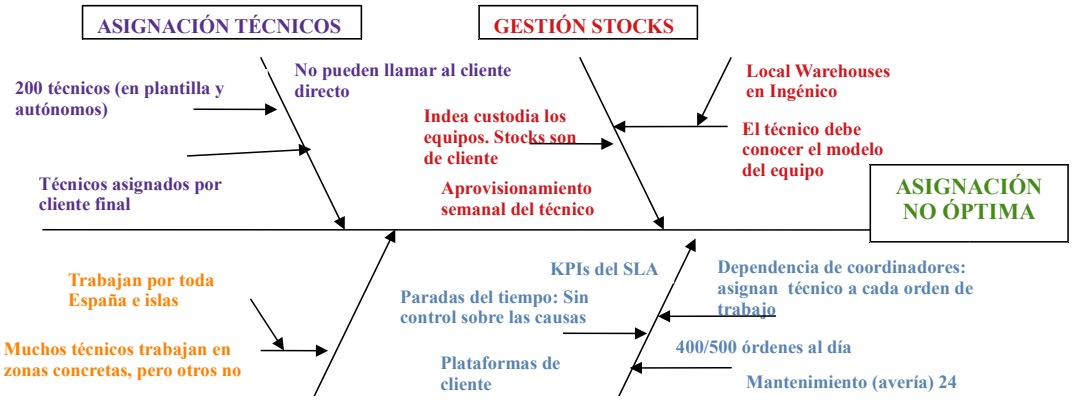


Ilustración 5: Diagrama de Ishikawa

Las principales fuentes consultadas ha sido la propia INDEA. Se ha trabajado principalmente con datasets facilitados por la empresa, resultados de exportaciones de campos acordados con el Director de las bases de datos de la compañía. Adicionalmente, también se ha obtenido información de SABI (Fundación EOI) e Internet.

- Brainstorming

  Sin duda, una de las técnicas principales usadas por el equipo de SonaR ha sido ésta, principalmente en la fase de concepción inicial del proyecto. Mediante esta técnica se ha

posibilitado a los miembros del grupo opinar sobre el problema a resolver, aprovechando así la capacidad creativa de los cuatro participantes del equipo de trabajo. Se han llevado a cabo numerosos brainstorming para preseleccionar las mejores ideas y dar forma al problema y el proyecto.

Durante la fase de identificación del problema, se observa que el proceso de asignación de recursos para la asignación de incidencias es claramente mejorable. La asignación de órdenes de trabajo a los técnicos ha de hacerse de acuerdo a las competencias requeridas y tiene en la actualidad una alta dependencia del criterio de cada coordinador. También ha de tener en cuenta zonas, horarios, experiencia, ubicación e incidencias que tiene abiertas en cada momento cada técnico, etc. Todas estas variables forman parte del proceso de asignación que actualmente se efectúa de forma manual por parte de coordinadores, de los cuales Indea se muestra completamente dependiente. Prueba de ello es el aumento de incumplimientos de SLA ante vacaciones o bajas puntuales de ciertos coordinadores, por una incapacidad de realizar el proceso de forma eficiente.

1.3. La oportunidad de negocio

Una vez identificado el problema que supone la asignación poco eficiente de los técnicos a las Órdenes de Trabajo, se detecta la oportunidad de negocio que supone optimizar este proceso. El proyecto SonaR busca **optimizar el proceso de asignación de Ordenes de Trabajo (OT) de INDEA**; por el volumen de Órdenes de Trabajo y técnicos, inicialmente se realizará un prototipo centrado en el cliente Grupo Ingénico y la provincia de Barcelona.

Como se ha comentado anteriormente, en la actualidad la asignación se realiza de forma directa por los coordinadores; este hecho provoca que la propia empresa INDEA haya identificado esta oportunidad de optimización y demande una solución que tenga en cuenta variables críticas, modelice el proceso de asignación y detecte anomalías, así como el aprendizaje de los técnicos. De este modo, el planteamiento de SonaR es dar respuesta a la necesidad de optimizar el proceso de asignación de OT, automatizando desde la ingesta de los datos, hasta una propuesta que visualizarán los trabajadores que actualmente realizan el proceso.

Para optimizar el proceso de asignación, se plantean técnicas de Inteligencia Artificial y Analítica sobre los datos disponibles. Para ello, se ha podido consensuar con la Dirección de Indea los datos necesarios de entre los disponibles. Se ha planteado el proyecto en base a la optimización del proceso de asignación de técnicos en base a sus capacidades, ejecutando las órdenes de trabajo dentro del límite temporal impuesto por el cliente y mejorando el servicio prestado. Los coordinadores de Indea,

que son las personas encargadas de asignar las órdenes de trabajo, verán apoyada su labor de asignación y podrán centrarse en el proceso de verificación gracias a la implantación del modelo.

SonaR supone una oportunidad ante la demanda de la empresa de automatizar el proceso, que asume un alto riesgo al recaer la responsabilidad de la asignación en personas, que sufren cansancio y existe un riesgo de movilidad laboral o ausencia. Si se cumple el objetivo de cumplir los SLA y mejorar el servicio, se podría extender el modelo a otros clientes. El proyecto resulta por tanto del máximo interés para Indea Indea; la viabilidad del proyecto ha sido posible gracias principalmente a tres factores principales:

- Cooperación de la empresa: los órganos directivos de Indea afirman que resolver los problemas descritos anteriormente es **una necesidad y una oportunidad** para su empresa**.** Se trata de una necesidad identificada hace aproximadamente tres años, para la cual no han sido capaces de obtener solución por sí mismos por falta de formación tecnológica y de capacidad operativa. Sin embargo, aunque no disponen de un equipo formado con tiempo para poder acometer un proyecto de esta magnitud, han reconocido en todo momento el valor de la optimización y transformación de la optimización del proceso para su negocio. Como consecuencia de lo anterior, desde la primera entrevista se cuenta con la máxima cooperación de los directivos de Indea, un trato excelente y las mayores facilidades para que el proyecto haya sido llevado a cabo, en cuanto a facilitar el acceso a la información y dar explicaciones del modelo de negocio.
- Disponibilidad de información en Indea y acceso a la misma por el equipo de SonaR: Indea ha venido recopilando información en relación al cliente Ingénico desde el año 2015, organizada en una base de datos *Firebird*. La información disponible se describe en detalle en el próximo apartado; cabe destacar en este momento únicamente la absoluta transparencia y colaboración que se ha obtenido en todo momento de la empresa para depurar y mejorar el dataset disponible.
- Volumetría adecuada para el alcance del proyecto: el volumen de datos disponibles es elevado, si bien manejable para el prototipo. Por ejemplo, para el periodo 1 de Enero a 26 de Noviembre de 2018 se dieron aproximadamente 104.000 órdenes de trabajo para el cliente Grupo Ingénico.
- Inversión necesaria: la inversión que ha resultado necesaria para llevar a cabo el prototipo consiste principalmente en el coste del equipo humano (horas del equipo de trabajo). Si bien se ha estimado este coste en el análisis económico detallado en el apartado 4, el desarrollo del piloto se ha efectuado gratuitamente para Indea, a modo de prueba de concepto y proyecto fin de Máster. Por su parte, en relación a la infraestructura utilizada, ha sido viable efectuar el entorno Amazon Work Spaces facilitado por la EOI como alumnos del Máster.

En conversaciones previas con los directivos de Indea se ha puesto de manifiesto que en la actualidad la ineficiencia del proceso y la dependencia con respecto a ciertos recursos está impidiendo escalar las operaciones, con lo que en caso de obtener resultados positivos en el prototipo con Grupo Ingénico, se vería con muy buenos ojos ampliar a un mayor número de clientes y de mayor relevancia, ya que pequeñas mejorías en los indicadores (un punto porcentual en el cumplimiento de SLAs) suponen importantes beneficios para la entidad por la elevada volumetría que maneja. Por tanto, no sería descartable que una vez demostrada la potencialidad del prototipo, será negociable una inversión por parte de la empresa para escalar la solución al resto de clientes y operaciones.

2. Investigación y análisis exploratorio de los datos

2.1. Modelo de datos

Para comenzar el análisis, Indea ha facilitado un conjunto de datos de las órdenes de trabajo manejadas para su cliente Ingénico, durante el año 2018 en todo el territorio nacional, resultando un total de 104.478 registros, en los que se incluyen ambos tipos de órdenes de trabajo, mantenimientos e instalaciones. Es importante destacar que:

- Los datos recibidos corresponden al periodo 1 de Enero de 2018 a 27 de Noviembre de 2018 (fecha de exportación de la información)
- Los datos incluyen las órdenes de trabajo de todos los clientes del grupo Ingénico, en los cuales INDEA debe realizar las órdenes de trabajo, con excepción de La Caixa, conjunto de datos que por volumetría y por contar con KPIs diferentes, requeriría de un análisis específico fuera del alcance acordado con la empresa para el proyecto.

El dataset contiene información de todas las provincias, si bien, como se mencionó en el primer capítulo, el piloto se centrará únicamente en una de ellas. En conversaciones con INDEA, se selecciona el subconjunto del dataset correspondiente a Barcelona, que es la provincia con mayor número de órdenes de trabajo, técnicos de campo y complejidad en la gestión.

En la tabla a continuación, se muestra la distribución de órdenes de trabajo en las 10 provincias con mayor volumetría:

Provincia	Número de órdenes de trabajo
BARCELONA	36434
MADRID	27981
BALEARES	4759
VALENCIA	4207
TARRAGONA	3984
GIRONA	3636
MURCIA	3405
ALICANTE	2825
GRANADA	1469
VALLADOLID	1090

Tabla 1. Número de órdenes de trabajo en las 10 provincias con mayor volumetría

El conjunto de datos recibido incluye los campos más relevantes para el modelo a juicio del Director de Sistemas de Información de INDEA y es válida para hacer un primer análisis de los datos. El conjunto de datos recibidos es el resultado de consulta a varias tablas de la base de datos Firebird del software PEZ, que como se mencionó en el capítulo anterior almacena toda la información del sistema.

En diagrama de base de datos relacional que se muestra a continuación, pueden observarse el conjunto de tablas definidas en la base de datos Firebird, sobre la cual basa INDEA su software de gestión de incidencias.

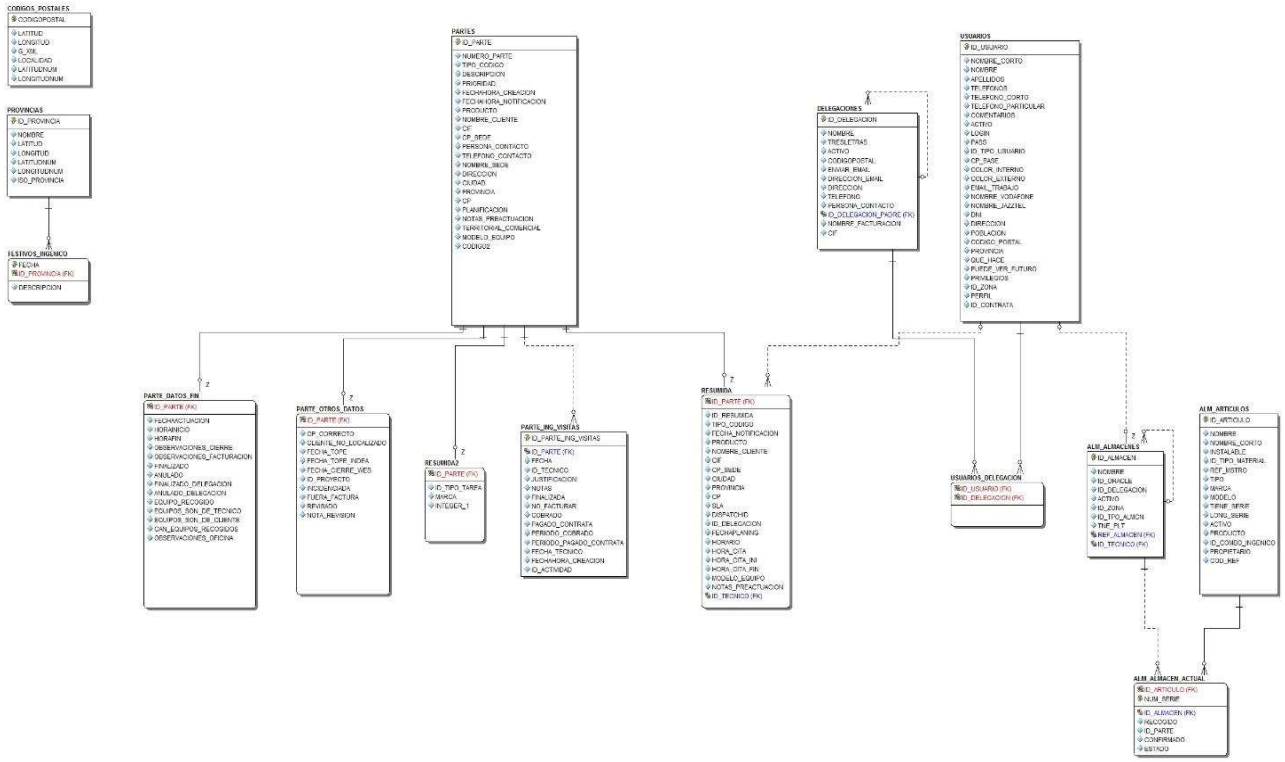


Ilustración 6. Diagrama de base de datos relacional Firebird de Indea

En resumen, el software de gestión PEZ dispone de los siguientes grandes bloques de información:

- Información de las órdenes de trabajo, incluyendo descripciones, fechas, modelos de equipo afectados, incidencias, datos sobre la actuación del técnico, estado, control de ANS, etc.
- Información sobre las citas de los técnicos con los usuarios finales
- Información sobre los usuarios finales
- Información sobre los almacenes
- Tablas auxiliares (códigos postales, provincias, festivos)

2.2. Descripción general de los datos

Para iniciar el análisis exploratorio, se utiliza como base el dataset exportado por la empresa INDEA. Los datos de los que se disponen son los siguientes:

CIUDAD	104439 non-null	object	Ciudad de la empresa cliente donde se instala el TPV
CLIENTE	104478 non-null	object	Nombre (corto) de la empresa cliente donde se instala el TPV
CLIENTE_INGENICO_LARGO	104478 non-null	object	Nombre (largo) de la empresa cliente donde se instala el TPV Contiene el nombre corto, un código unívoco e información adicional, todo ello separado por ;
CP	104441 non-null	object	Código postal de la empresa cliente
ESTADO_EVENTO	104478 non-null	object	Estado con el que está catalogada el Evento: Resuelta, Anulada, Sin Resolver
FEC_TOPE_MENOS_FEC_RESUELTA	104478 non-null	object	Tiempo resultante de restar a la fecha límite para la resolución del Evento la fecha en la que se resuelve
FH_ENTRADA_EVENTO.Date	104478 non-null	datetime64[ns]	Día en el que entra el Evento en el Sistema
FH_PLANIFICADA_EVENTO.Date	104478 non-null	object	Día en el que se asigna el Evento
FH_REALIZADA_TECNICO_EVENTO.Date	104478 non-null	object	Día en el que el Técnico realiza la actuación
FH_RESUELTA_CESTRACK_EVENTO.Date	104478 non-null	object	Día en el que se resuelve el Evento en el Sistema
FH_TOPE_EVENTO.Date	104478 non-null	object	Día límite para la resolución del Evento
HORAS_PARA_CUMPLIR_SLA	104478 non-null	int64	Horas restantes hasta llegar a la fecha límite de resolución del Evento
ID_EVENTO	104478 non-null	object	Número con el que se identifica el Evento
ID_TECNICO	104478 non-null	object	Número identificativo de cada Técnico
INCIDENCIADO	104478 non-null	object	Se identifica con (0,1) los Eventos incidenciados
LATITUD	104478 non-null	object	Latitud, coordenadas de la empresa cliente
LONGITUD	104478 non-null	object	Longitud, coordenadas de la empresa cliente

NIS	104478 non-null	object	Se valora como 0 un Evento sin incidencias y como 1 cuando el Evento tiene incidencia
NIS_MOTIVO	24148 non-null	object	Motivo de la incidencia
NOMBRE_CORTO_TECNICO	104478 non-null	object	Nombre (corto) del Técnico asignado
PREV_EQUIPOS_INSTALAR_CAMBIAR	104478 non-null	object	Nombre del dispositivo que va a necesitar para el Evento
PROVINCIA	104478 non-null	object	Provincia donde se encuentra la empresa cliente
SLA_OK	104478 non-null	object	Se ha resuelto el Evento antes de la fecha límite (Si/No)
SLA_OK_INCIDENCIA	104478 non-null	object	Se ha resulto el Evento tras la incidencia antes de la fecha límite (Si/No)
TECNICO	104478 non-null	object	Nombre completo del Técnico asignado
TIPO_CODIGO_EVENTO	104478 non-null	object	Código con el que se identifica el tipo de Evento realizado
TIPO_EVENTO	104478 non-null	object	Se identifican los Eventos como M/I en función del tipo de actuación

Tabla 2. Variables contenidas en el dataset proporcionado por Indea

En la tabla anterior, se puede observar, en primer lugar, que la mayoría de las variables son de tipo "object" y la existencia de campos redundantes y valores nulos. Por ello, es necesario realizar un proceso de Análisis Exploratorio de los Datos (EDA), como parte previa al ETL previo al entrenamiento del modelo predictivo con el que se pretende dar solución al problema planteado en el proyecto.

2.3. Conjunto de datos. Planteamiento de EDA

La analítica exploratoria (*Exploratory Data Analytics* EDA) es un conjunto de tareas que tienen como objetivo entender y preparar el dataset de datos de trabajo para una posterior aplicación de técnicas de modelado y *machine learning.* Estas técnicas son de naturaleza gráfica y cuantitativa.

Una vez analizada la información preliminar, se filtra el dataset para obtener únicamente los registros correspondientes a la provincia de Barcelona, obteniéndose el siguiente dataset, compuesto por 35.365 registros y 27 columnas:

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 35365 entries, 0 to 35364
Data columns (total 27 columns):
CIUDAD                              35360 non-null object
CLIENTE                             35365 non-null object
CLIENTE_INGENICO_LARGO              35365 non-null object
CP                                  35365 non-null int64
ESTADO_EVENTO                       35365 non-null object
FEC_TOPE_MENOS_FEC_RESUELTA         34950 non-null float64
FH_ENTRADA_EVENTO                   35365 non-null object
FH_PLANIFICADA_EVENTO               35326 non-null object
FH_REALIZADA_TECNICO_EVENTO         34959 non-null object
FH_RESUELTA_CESTRACK_EVENTO         34950 non-null object
FH_TOPE_EVENTO                      35365 non-null object
HORAS_PARA_CUMPLIR_SLA              35365 non-null int64
ID_EVENTO                           35365 non-null object
ID_TECNICO                          35365 non-null object
INCIDENCIADO                        35365 non-null object
LATITUD                             35365 non-null object
LONGITUD                            35365 non-null object
NIS                                 35365 non-null object
NIS_MOTIVO                          4908 non-null object
NOMBRE_CORTO_TECNICO                35365 non-null object
PREV_EQUIPOS_INSTALAR_CAMBIAR       35365 non-null object
PROVINCIA                           35365 non-null object
SLA_OK                              35365 non-null object
SLA_OK_INCIDENCIA                   35365 non-null object
TECNICO                             35365 non-null object
TIPO_CODIGO_EVENTO                  35365 non-null object
TIPO_EVENTO                         35365 non-null object
dtypes: float64(1), int64(2), object(24)
memory usage: 7.3+ MB
```

Ilustración 7. Información resumida de los datos de partida para el desarrollo del proyecto

Una primera descripción de la información contenida en el dataset permite ver que casi todos los campos han sido interpretados como de tipo *object,* lo cual imposibilita el entrenamiento de modelos, siendo preciso realizar:

- Análisis de las diferentes variables que componen el dataset con el objetivo de optimizar el conjunto de entrenamiento, utilizando principalmente técnicas de visualización y correlación.
- Carga, transformación y limpieza de variables
- Creación de nuevas variables identificados como necesarios para el entrenamiento del modelo, después de las diferentes encuestas realizadas a la empresa

2.4. Análisis de variables

Tipo de campo, descripción, naturaleza, cardinalidad

En el apartado a continuación, se describe el tipo de dato de cada variable y los niveles de información que contiene. Para ello, se parte del dataset global, no el específico para el caso piloto de Barcelona, dado que se persigue tener una visión general de los campos que permita extraer conclusiones generales y permitan el crecimiento del prototipo piloto hacia futuras expansiones de la herramienta:

- ID_EVENTO: Cadena de caracteres de longitud 10 que identifica unívocamente una incidencia
- PROVINCIA, CIUDAD, CP: Existen 3 campos de caracteres de longitud variable que tienen relación con el nombre o código de la localización de la incidencia.
    - PROVINCIA: tiene 53 levels (las 52 provincias y "-"). Hay valores no válidos identificados con el caracter "-" que debería eliminarse o deducir del valor de CIUDAD. Este campo, aunque se mantendrá para el análisis, se eliminarás posteriormente en el entrenamiento del modelo, dado a que el prototipo objeto de este proyecto, estará basado únicamente en una provincia.
    - CIUDAD tiene 5114 levels. algunos inválidos con valos nulo ("") y otros con inconsistencias en el string correspondiente al valor ("A COROÑA", "A CORUNA", "A CORUÑA"), es por esta razón que se eliminará este campo para el posterior análisis.
    - CP es el código postal del local de la empresa donde se realiza la intervención. Tiene valores no válidos "" y "." en el dataset, estos valores serán eliminados posteriormente al hacer el análisis de valores nulos.
- LATITUD: Corresponde a la latitud geográfica de las coordenadas de la empresa donde se realiza la intervención.
- LONGITUD: Corresponde a la longitud geográfica de las coordenadas de la empresa donde se realiza la intervención
- CLIENTE: Campo de caracteres de longitud variable que identifica al cliente en el que se debe resolver la incidencia.
- FH_ (CAMPOS de Fecha en formato Datetime): FH_ENTRADA_EVENTO, FH_PLANIFICADA_EVENTO, FH_REALIZADA_TECNICO_EVENTO, FH_RESUELTA_CESTRACK_EVENTO, FH_TOPE_EVENTO
- FEC_TOPE_MENOS_FEC_RESUELTA: Variable numérica que proporciona información redundante. En principio se considerará su eliminación o sustitución en el proceso ETL.

- HORAS_PARA_CUMPLIR_SLA: Variable numérica que indica el número de horas para cumplir el SLA. En el dataset disponible, aparecen valores negativos cuando se ha excedido el número de horas (incumplimiento de SLA)
- ESTADO_EVENTO: Cadena de caracteres con 9 niveles. Los valores proporcionan información de si la incidencia ha sido resuelta, está anulada o sin resolver: [1] "ANULADO PEZ"; [2] "RESUELTO CESTRACK EN SLA"; [3] "RESUELTO CESTRACK NO SLA"; [4] "RESUELTO PEZ EN SLA"; [5] "RESUELTO PEZ NO SLA"; [6] "RESUELTO TÉCNICO"; [7] "SIN RESOLVER"; [8] "SIN RESOLVER INCIDENCIADO: KO Falta Adecuación Línea"; [9] "SIN RESOLVER INCIDENCIADO: KO No quiere TPV"
- ID_TECNICO: Cadena de caracteres numéricos (182 levels) que representan unívocamente cada técnico. Existen valores con "-" que habrá que limpiar para el análisis.
- PREV_EQUIPOS_INSTALAR_CAMBIAR: Cadena de caracteres indicando información del equipo a instalas. La cadena arrastra retorno de carro final de la cadena. El campo tiene 109 valores.
- INCIDENCIADO: Valor que indica que la intervención tiene una incidencia. Valores: "NO" (100.183 casos) "SI" (4309 casos). Los equipos más incidenciados son:

	Var1 <fctr>	Freq <int>	percent <dbl>
20	ICT250 MIXTO-GPRS_x000D_\n	816	0.2059046177
41	IWL288_x000D_\n	644	0.1625031542
37	IWL281 GPRS CL_x000D_\n	602	0.1519051224
44	MOVE 5000 GPRS_x000D_\n	350	0.0883169316
36	IWL251_x000D_\n	222	0.0560181681
21	ICT250 S CL_x000D_\n	122	0.0307847590
68	VERIFONE VX680-WIFI_x000D_\n	116	0.0292707545
71	VX680-WIFI_x000D_\n	102	0.0257380772
40	IWL288 WIFI CTLES_x000D_\n	101	0.0254857431
33	IWL222_x000D_\n	91	0.0229624022

Tabla 3. Equipos con mayor número de incidencias

- NIS: Campo alfanumérico de 3 niveles con valores "0", "1" y "-" relacionado con la existencia de incidencia en la intervención
- NIS_MOTIVO: Cadena de caracteres describiendo brevemente el motivo de la incidencia ("Ausencia Cliente", "Dificultad de Contacto", "Falta autorizacion cliente", "Falta Cobertura", ect). Podemos ver la combinación de Técnico – Cliente que presenta un como motivo de incidencia la "Dificultad de Contacto"

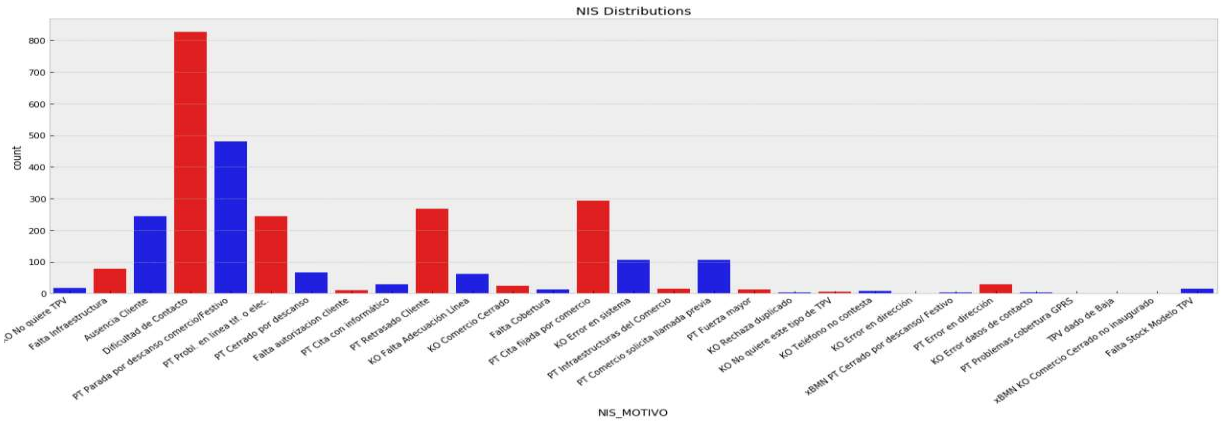


Ilustración 8. Distribución de los motivos de las incidencias

- SLA_OK: SI (si se ha cumplido) / NO (no se ha cumplido o la intervención está en curso)
- SLA_OK_INCIDENCIA
    - SI: SLA se ha cumplido con incidencia resuelta
    - NO: SLA no se ha cumplido y el caso está incidenciado
- TIPO_CODIGO_EVENTO: Cadena de 5 caracteres con los valores "ING-C" "ING-I" "ING-M" "ING-R". Indica si es nueva instalación, mantenimiento u otros. Probablemente información redundante
- TIPO EVENTO: Valores "I" "M", indicando instalación y mantenimiento

Análisis de datos para el proyecto piloto

En los subapartados anteriores se ha analizado la información del dataset de Indea para todas las provincias. De forma análoga, se ha realizado un análisis equivalente del comportamiento de las variables para el subconjunto de datos del piloto (ordenes de trabajo en la provincia de Barcelona), que permiten realizar un proceso ETL más enfocado al problema específico. En el presente apartado no se replican todas las explicaciones incluidas para el proceso general, sino que se recogen exclusivamente algunas conclusiones relevantes del análisis efectuado.

- En el conjunto de los datos hay 30.739 casos que han cumplido el SLA (86,92%), mientras que 4.626 no cumplen el SLA (13,08%), según se muestra en la siguiente figura:

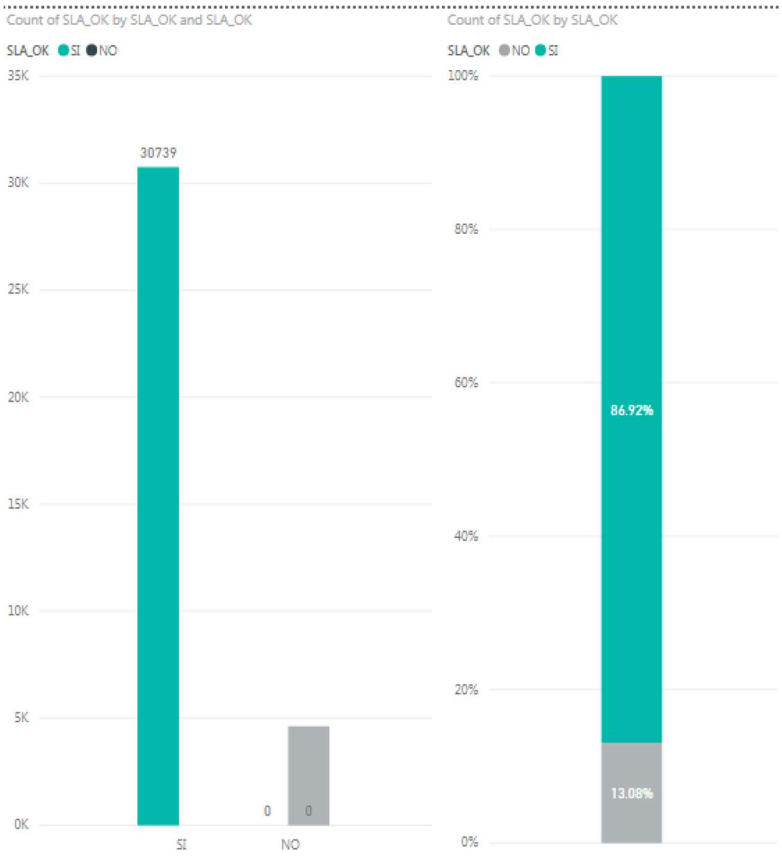


Ilustración 9. Distribución de cumplimiento de SLA

- Analizando a todos los técnicos, observamos que en términos generales el cumplimiento se sitúa por encima de 87% de las tareas, según se muestra en la figura siguiente:

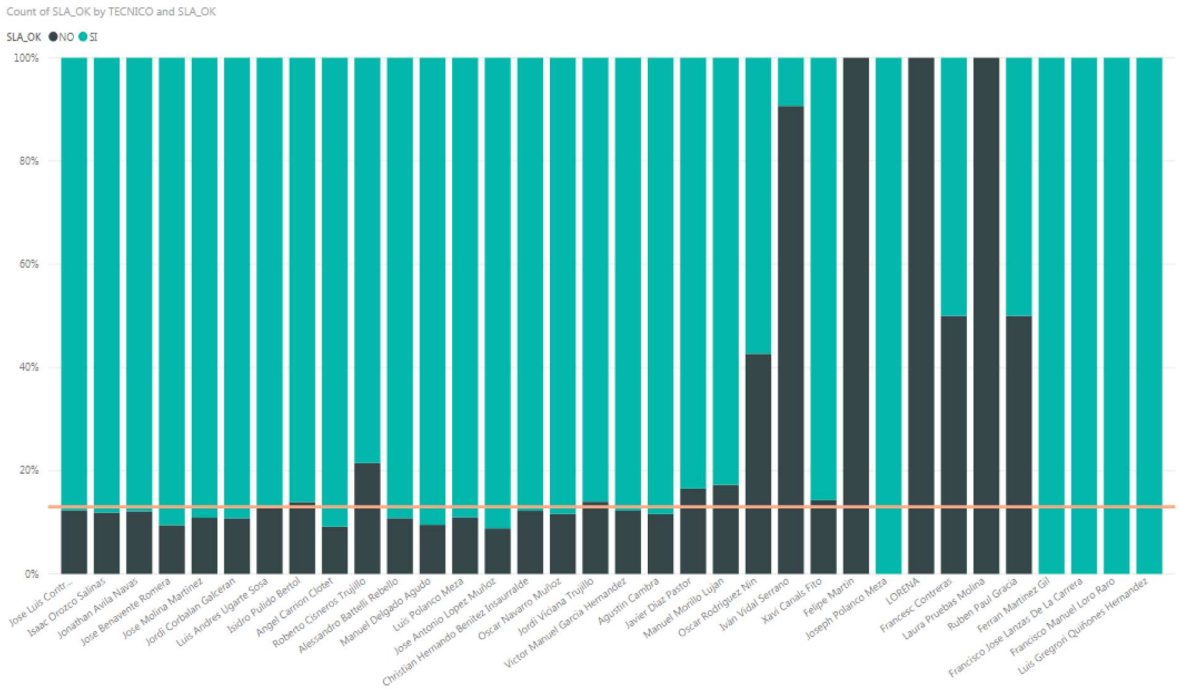


Ilustración 10. Distribución de cumplimiento de SLA por técnico en Barcelona

- CHIPCARD incumple en un 69,14% de las tareas que se han realizado para ellos (746 tareas), el 54,77% de los casos que no cumplen con esta marca equivalen a una instalación del modelo USB-GST-1252-C1 (591 tareas), y el 11,4% al modelo USB-GST-1252-G1 (123 tareas).
- Ikea-Telefónica incumple en un 44,23% de los casos (23 tareas), mientras que cumplen en 29 casos. En todos los casos donde no cumplieron SLA se instaló el modelo ISC-350.
- VIPS incumple en un 43,48% de las instalaciones (50 tareas), mientras que 65 tareas si cumplen con el SLA. En este caso el 29,57% de los casos que incumplieron no tienen aparato asociado a la tarea (34 casos), como se puede comprobar en la siguiente figura:

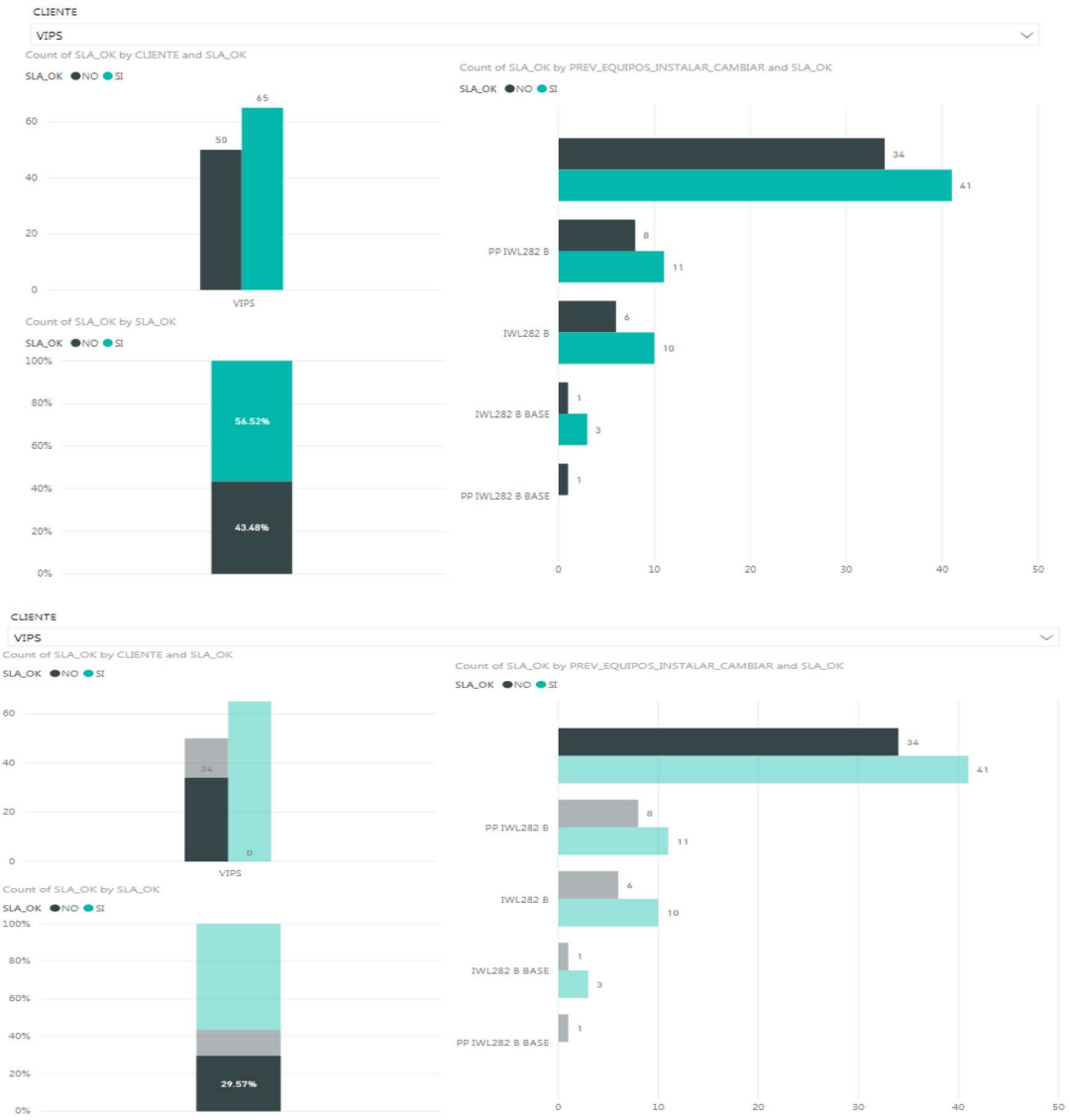


Ilustración 11. Cumplimiento de SLA en cliente VIPS

- El modelo USB-GST-1252-C1 acumula 591 actuaciones en las que no cumple el SLA, todas ellas con el cliente CHIPCARD. Los técnicos que más veces han tenido una tarea que haya incumplido el SLA con este modelo son R. C. (164 tareas, equivalen al 19,57% de las incidencias relacionadas con este modelo), I.P. (94 casos incumplidos, equivalen al 11,22% de los casos), L.A.U. (67 casos incumplidos, el 8% de los casos totales), J.L.C.M. (49 casos incumplidos, que equivalen al 5,85%), J.A. (48 casos incumplidos, equivalentes al 5,73%) e I.O.

(42 casos incumplidos, un 5,01% del total). En total este modelo tiene un total del 70,53% de casos en los que no ha cumplido el SLA (véase figura siguiente).

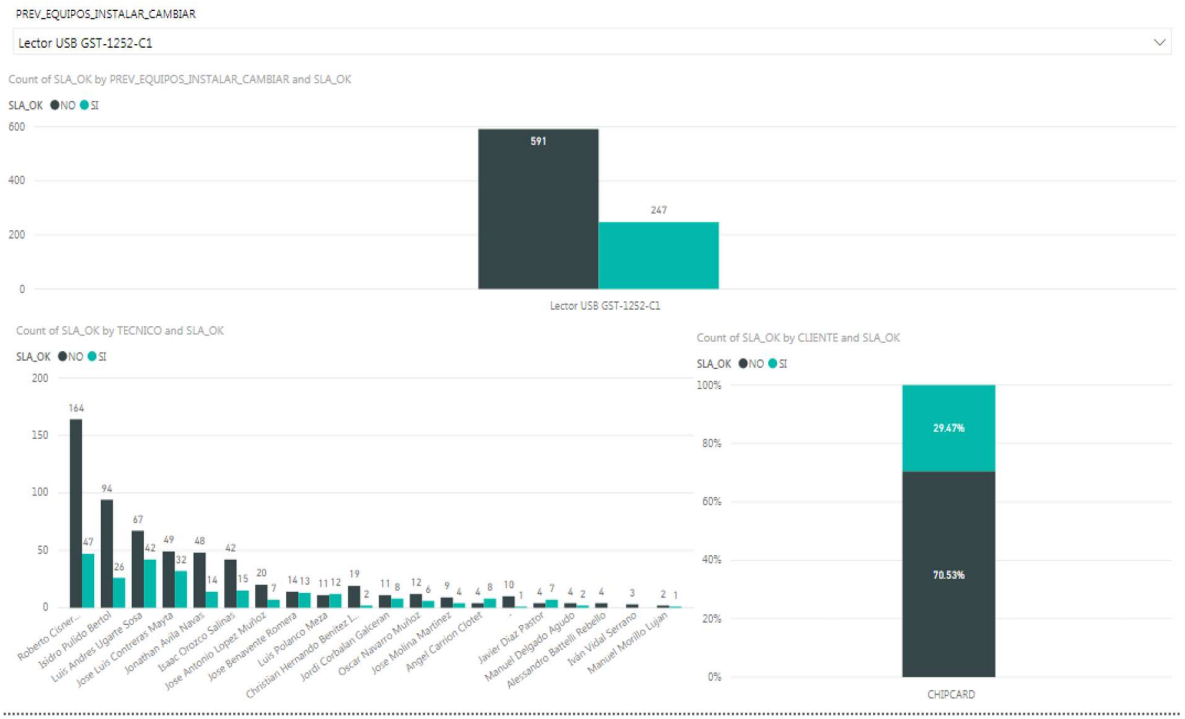


Ilustración 12. Análisis de cumplimiento de SLA para el equipo USB-GST-1252-C1

- El modelo ISC-350 tiene un 46% de los casos con SLA KO, 11 tareas realizadas por R. C., 8 tareas por J. M., 2 por C. H. y 1 tarea realizada por J. C. Todos ellos cuentan con un 100% de KO en instalaciones con este modelo, mientras que los demás casos con SLA OK (27 tareas), los técnicos tienen un 0% de incumplimiento en sus tareas realizadas.

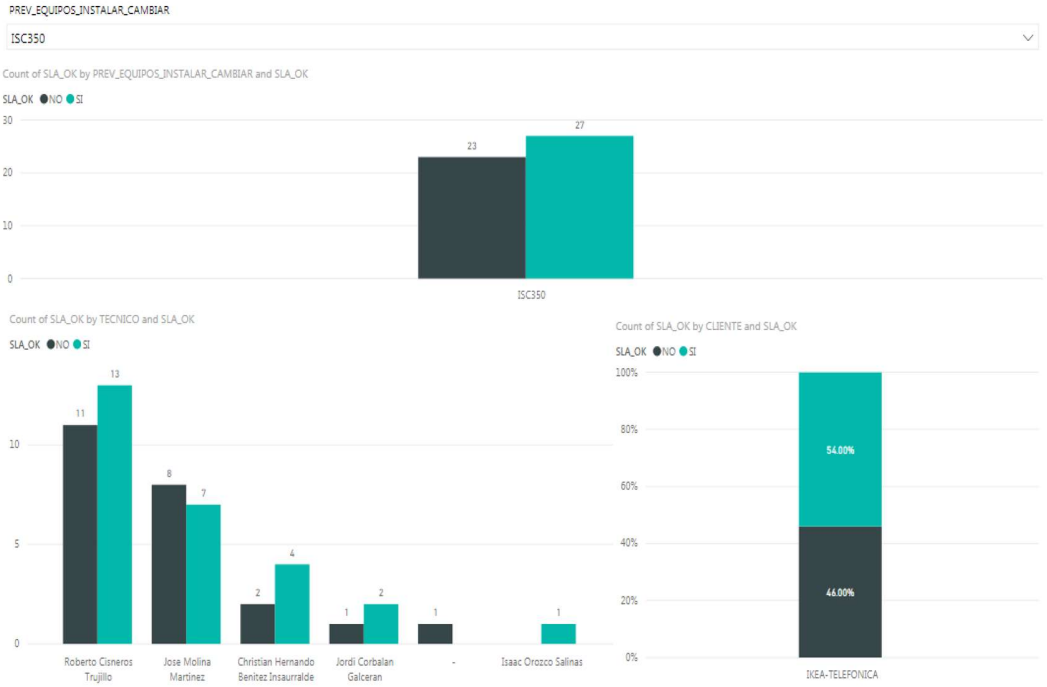


Ilustración 13. Cumplimiento de SLA para el equipo ISC-350

2.5. Proceso ETL

Una vez analizadas las variables, y teniendo en consideración lo expuesto para cada una de ellas, a continuación, se listan los pasos que se han efectuado como parte del proceso ETL para la optimización del conjunto de datos:

Tratamiento de valores inesperados

El primer paso consiste en la eliminación de valores inesperados, ya que estos valores imposibilitan realizar transformaciones sobre los campos (por ejemplo, no se puede convertir a entero una columna).

Tras el análisis efectuado sobre el dataset, y tal como se ha adelantado en la descripción del análisis de algunas variables, se detecta que en algunas variables aparecen campos con caracteres"-" cuando no se conoce el dato, por lo cual, en tales casos, se procede a la sustitución de los caracteres "-" por valores nulos.

Adicionalmente, se ha podido identificar la presencia de caracteres \r\n (CR LF) en los valores de la columna PREV_EQUIPOS_INSTALAR_CAMBIAR, por lo que se procede a eliminar dichos caracteres.

Transformación de tipos de variables

En primer lugar, se procede a la transformación de variables que contienen valores numéricos pero han sido identificadas como strings; tal es el caso de NIS, LATITUD o LONGITUD. En el caso de la LATITUD y la LONGITUD, es necesario adicionalmente transformar las comas en puntos.

Por su parte, los campos de FH_ENTRADA_EVENTO, FH_PLANIFICADA_EVENTO, FH_REALIZADA_TECNICO_EVENTO, FH_RESUELTA_CESTRACK_EVENTO y FH_TOPE_EVENTO se transforman al tipo datetime. Cabe destacar que se trata de formato fecha y hora y no únicamente fecha, por el tipo de operaciones a efectuar en el transcurso del proceso ETL.

Finalmente, se aplican las dos técnicas de linealización más frecuentes:

- *Label Enconding:* La transformación del tipo de algunas variables categóricas, que admiten un número limitado de valores textuales, a valores numéricos es inmediata; de este modo:
    - TIPO_EVENTO (1 si es I y 2 si es M)
    - SLA_OK (1 si es SI y 0 si es NO)
    - INCIDENCIADO (1 si es SI y 0 si es NO)
    - SLA_OK_INCIDENCIA (1 si es SI y 0 si es NO)
- *One-Hot Encoding:* se procede a linealizar las variables categóricas presentes en el dataset, entre las variables que se linealizan se encuentran: 'CLIENTE', 'NIS_MOTIVO', 'PREV_EQUIPOS_INSTALAR_CAMBIAR' y 'ID_TECNICO'. Con esta operación, el número de operaciones se incrementa sustancialmente, pasándose de un total inicial de 20 columnas a un total de 153 columnas. Las variables 'CLIENTE', 'NIS_MOTIVO', 'ID_TECNICO' y 'PREV_EQUIPOS_INSTALAR_CAMBIAR' se linealizan con la finalidad de conocer más adelante la relación de cada uno de los posibles valores que toman estas variables en el cumplimiento del SLA.

Tratamiento de valores nulos y registros no válidos

Durante el proceso de observación del dataset, se pudo identificar la presencia de algunos valores nulos o cero en los datos. En primer lugar, se procede a inspeccionar los valores nulos existentes en la tabla para cada una de las columnas:

CIUDAD	5
CLIENTE	0
CLIENTE_INGENICO_LARGO	0
CP	0
ESTADO_EVENTO	0
FEC_TOPE_MENOS_FEC_RESUELTA	415
FH_ENTRADA_EVENTO	0
FH_PLANIFICADA_EVENTO	39
FH_REALIZADA_TECNICO_EVENTO	406
FH_RESUELTA_CESTRACK_EVENTO	415
FH_TOPE_EVENTO	0
HORAS_PARA_CUMPLIR_SLA	0
ID_EVENTO	0
ID_TECNICO	285
INCIDENCIADO	0
LATITUD	25
LONGITUD	25
NIS	412
NIS_MOTIVO	30869
NOMBRE_CORTO_TECNICO	285
PREV_EQUIPOS_INSTALAR_CAMBIAR	0
PROVINCIA	0
SLA_OK	0
SLA_OK_INCIDENCIA	0
TECNICO	285
TIPO_CODIGO_EVENTO	0
TIPO_EVENTO	0

De lo anterior se puede observar, que en el caso de la columna **NIS_MOTIVO** existen muchos valores nulos, pero como esto se debe a que esta columna solo tiene información en caso de que el ticket haya presentado alguna incidencia durante su resolución (es decir, que el valor de NIS sea igual a 1), se procederá a reemplazar estos valores nulos por el string **NIS igual a 0** en el caso en el que el valor nulo se deba a que el NIS es igual a cero, y por el string **"Sin Especificar"** en el caso en el que sea un valor nulo para casos con NIS igual a uno. A continuación, se procede a eliminar todos los registros con valores nulos identificados con la finalidad de limpiar el dataset. Se observa en esta limpieza que:

- El dataset se reduce a 34.884 registros (481 registros menos que los iniciales 35.365 registros), el cual sigue siendo una información suficientemente voluminosa para afrontar el entrenamiento del modelo.
- Hay un total de 39 valores nulos en el campo CIUDAD y un total de 37 nulos en el campo CP. Esta información quizá pudiera ser completada a partir de datos de LATITUD/LONGITUD, si bien se ha desestimado por suponer un número despreciable de registros (frente a los más de 35.000 datos disponibles).

Adicionalmente, existen distintos estados en el campo ESTADO_EVENTO que corresponden con partes de trabajo anulados (ANULADO) o no terminados (se exportó la base de datos en un momento en el que no se habían finalizado); se procede a eliminar los registros en esos estados, manteniendo únicamente la información relacionado con el estado RESUELTO EN CESTRACK.

Operaciones para garantizar la coherencia de los datos

Se procede a realizar una serie de operaciones que permitan garantizar la calidad de los datos que se usarán posteriormente para entrenar el modelo, comprobando la coherencia de los siguientes datos:

- Si el campo FEC_TOPE_MENOS_FEC_RESUELTA muestra valores negativos, lo cual implica incumplimiento del SLA, el valor del campo SLA_OK debería ser siempre 0.
- Si un ticket tiene el campo INCIDENCIADO igual a 1, el campo NIS debería ser igual a 1 y el campo NIS_MOTIVO debería ser diferente de "Sin Especificar"
- Se comprueba que el campo FH_REALIZADA_TECNICO_EVENTO nunca puede ser mayor que el campo FH_RESUELTA_CESTRACK_EVENTO

Se procede a eliminar los registros incoherentes y se comprueba que todos los tickets con incidencias hayan sido correctamente eliminados, no existiendo incoherencias en ese punto.

Eliminación de campos

Por las conversaciones mantenidas con la empresa y su conocimiento de la lógica de negocio, es evidente que algunos campos no aportan información, al estar contenidos en otros o no tener relevancia para el posterior entrenamiento del modelo, procediéndose a su eliminación:

Campos geográficos:

- Se elimina PROVINCIA: provincia de la incidencia. Al realizarse un piloto con BARCELONA, toma el mismo valor para todos los registros.
- Se elimina CIUDAD, ya que la información del CP contiene intrínsecamente esa información

Campos relacionados con los agentes involucrados:

- Se eliminan TECNICO y NOMBRE_CORTO_TECNICO, ya que contienen información textual de identificación del técnico que realiza la instalación o mantenimiento, previsiblemente no aportan información existiendo ID_TECNICO
- Se elimina CLIENTE_INGENICO_LARGO y se mantiene únicamente CLIENTE, ya que contiene la misma información que el primero

Otros campos:

- Se elimina ID_EVENTO, se trata de un identificador correlativo, unívoco
- Se eliminan ID_EVENTO y TIPO_CODIGO_EVENTO, por no aportar información relevante para el entrenamiento del modelo

2.6. Creación de nuevos campos

Según conversaciones con la empresa, los coordinadores, a la hora de realizar la asignación de incidencias a los técnicos en el actual proceso manual, consideran relevante cierta información que no aparece de forma explícita en el dataset. Se ha procedido a crear nuevos campos que recojan esa información:

- **Tiempo que tarda un técnico en resolver una incidencia:** Se obtiene de forma inmediata a partir de la resta entre dos campos tipo *datetime* disponibles: hora en la que concluye la instalación el técnico menos hora en la que entra el evento. Esta diferencia es una duración (tipo *timedelta*) se puede convertir de forma sencilla en entero (redondeando el número de horas, sin ser necesario discernir a nivel de minutos / segundos).
- **Distancia recorrida para resolver una incidencia (en kilómetros):** Este valor se calcula a partir de la longitud y latitud del cliente donde se ubica la incidencia, realizando la transformación geográfica correspondiente. Cabe destacar que para cada día, la primera distancia se calcula desde el punto de partida de cada técnico, y una vez finalizada la primera incidencia, se va calculando la distancia con respecto al punto de la incidencia anterior.
- **Número de incidencias pendientes de resolución mientras se está asignando una nueva incidencia:** Este valor se obtiene a partir del conteo de las incidencias cuya fecha y hora de apertura en el sistema era anterior y su fecha y hora de resolución era posterior a la fecha de la incidencia que se está asignando en ese momento.
- **Experiencia del técnico ante una orden de trabajo:** Se realiza una estimación a partir de la experiencia de un técnico en un cliente concreto, con el modelo de equipamiento concreto con el que trabajar, y su grado de cumplimiento general de SLA. Se efectúa en cuatro pasos:

  a) cálculo de la matriz TECNICO - CLIENTE - EXPERIENCIA en cliente (número de veces que ha visitado un cliente)

  b) cálculo de la matriz TECNICO - EQUIPO - EXPERIENCIA en equipo (número de veces que ha arreglado cada tipo de equipo)

  c) cálculo de la matriz con el grado de cumplimiento del SLA de cada técnico

  d) cálculo de la experiencia como una combinación a partir de los dos datos anteriores. Inicialmente se premia el cumplimiento de SLAs por parte del técnico como principal medida de su experiencia, aplicando la fórmula :

  *100 * experiencia del técnico en equipo + 100 * experiencia del técnico en un cliente + 1 * experiencia del técnico en cumplimiento en SLA*

si bien se realizarán las adaptaciones oportunas según se entrene el modelo.

- **Experiencia general del técnico:** Se realiza una estimación a partir de la experiencia de un técnico en el total de clientes, en el total de equipos existentes, y su grado de cumplimiento general de SLA. Se efectúa en cuatro pasos:

  a) cálculo de la matriz TECNICO - CLIENTE - EXPERIENCIA en cliente (número de veces que ha visitado un cliente)

  b) cálculo de la matriz TECNICO - EQUIPO - EXPERIENCIA en equipo (número de veces que ha arreglado cada tipo de equipo)

  c) cálculo de la matriz con el grado de cumplimiento del SLA de cada técnico

  d) cálculo de la experiencia como una combinación a partir de los dos datos anteriores. Inicialmente se premia el cumplimiento de SLAs por parte del técnico como principal medida de su experiencia, aplicando la fórmula :

  *100 * experiencia del técnico en equipos + 100 * experiencia del técnico en clientes + 1 * experiencia del técnico en cumplimiento en SLA*

Cabe destacar que la estimación de los diferentes componentes de la experiencia parte de la simplificación de que la experiencia de un técnico es constante en un año (el 2018).

En el planteamiento inicial se identificaron campos adicionales, que se han descartado al no disponer de información en el dataset para efectuar su cálculo; entre ellos, destacar los relacionados con el stock de equipos en el momento del alta de cada incidencia, la disponibilidad del técnico y el horario del comercio, entre otros.

2.7. Otras operaciones del proceso EDA

Análisis de Variables

En el siguiente apartado, se procede a realizar un análisis de las variables presentes en el dataset, con la finalidad de identificar cuál de ellas puede tener más impacto en la creación del modelo.

Para ello, en primer lugar, se usa un histograma para verificar si el campo FH_ENTRADA_EVENTO está relacionado de alguna manera con el no cumplimiento del SLA.

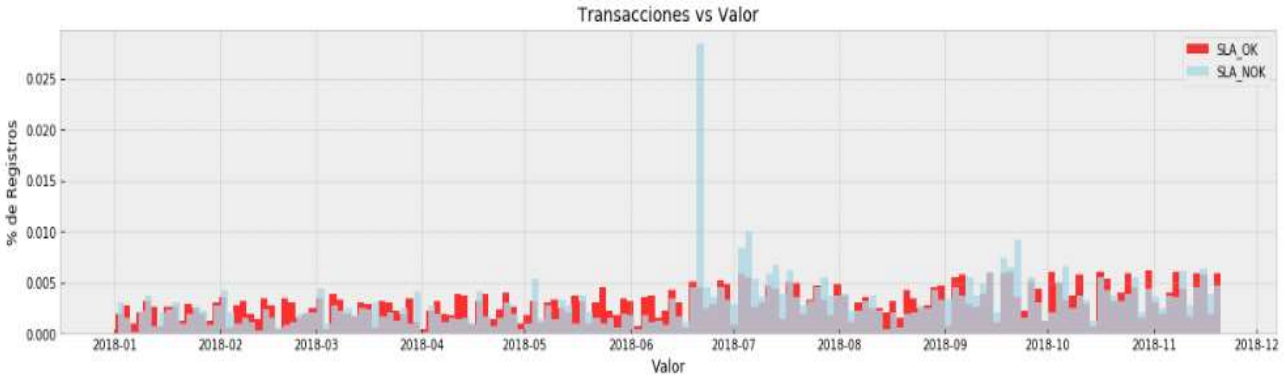


Ilustración 14. Relación (histograma) de la variable FH_ENTRADA_EVENTO con el cumplimiento de SLA

La gráfica anterior muestra claramente que, durante el periodo del verano, el incumplimiento del SLA sube, en principio bajo la sospecha que hay menos recursos humanos para atender el volumen de incidencias.

Sin embargo, también se observan algunas tendencias a la subida del no cumplimiento del SLA en algunos periodos de otros meses del año, razón por la cual se decide crear tres nuevas columnas en el dataset a partir del campo FH_ENTRADA_EVENTO: DIA SEMANA, DIA MES y MES.

Estos campos se crean con la finalidad de analizar si existe alguna tendencia de incumplimiento de SLA relacionada con la fecha de entrada del evento. Se procede a representar mediante histogramas la relación existente entre las nuevas variables creadas y el cumplimiento del SLA, resultando las siguientes gráficas:

Ilustración 15. Relación (histograma) de la variable WEEK_DAY con el cumplimiento de SLA

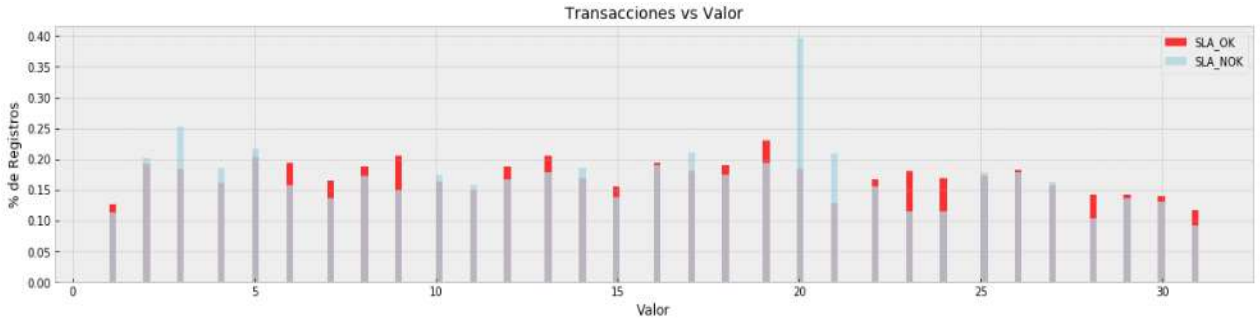


Ilustración 16. Relación (histograma) de la variable MONTH_DAY con el cumplimiento de SLA

Ilustración 17. Relación (histograma) de la variable MONTH con el cumplimiento de SLA

De lo anterior, se puede concluir, que durante los días viernes y sábados, la tendencia de no cumplir SLA crece, así como también durante los meses de verano.

En el anexo identificado como **Notebook - SonaR - ETL, EDA y Creacion de Variables.ipynb**, se muestra un análisis de la distribución de cada una de las variables del dataset en relación al cumplimiento del SLA. Del análisis realizado en el anexo anterior, se decide eliminar las siguientes variables, debido a que presentan una distribución uniforme en el cumplimiento del SLA: 'CP_8007', 'CP_8009', 'CP_8010', 'CP_8181', 'CP_8186', 'CP_8195', 'CP_8214', 'CP_8222', 'CP_8226', 'CP_8241', 'CP_8348', 'CP_8393', 'CP_8397', 'CP_8415', 'CP_8729', 'CP_8730', 'CP_8734', 'CP_8739', 'CP_8757', 'CP_8759', 'CP_8780', 'CP_8792', 'CP_8811', 'CP_8918', 'CP_8922', 'INCIDENCIAS_ABIERTAS', 'LATITUD', 'LONGITUD'.

Posteriormente, se procede a identificar las variables que hay que normalizar, que serán aquellas cuyo valor mínimo sea menor que -1 y el malor máximo mayor que 1, concluyéndose que la única variable que necesita normalización es **FEC_TOPE_MENOS_FEC_RESUELTA**

Correlación entre variables

El mapa de correlación entre las variables resultantes tras el proceso de ETL es el siguiente. Se presenta a alto nivel para extraer conclusiones, se puede consultar a mayor resolución en el notebook **SonaR - ETL, EDA y Creacion de Variables.ipynb**). En él se puede comprobar que el proceso ETL ha eliminado las correlaciones que existían en el conjunto inicial de datos, observándose que apenas existe correlación en el nuevo dataset.

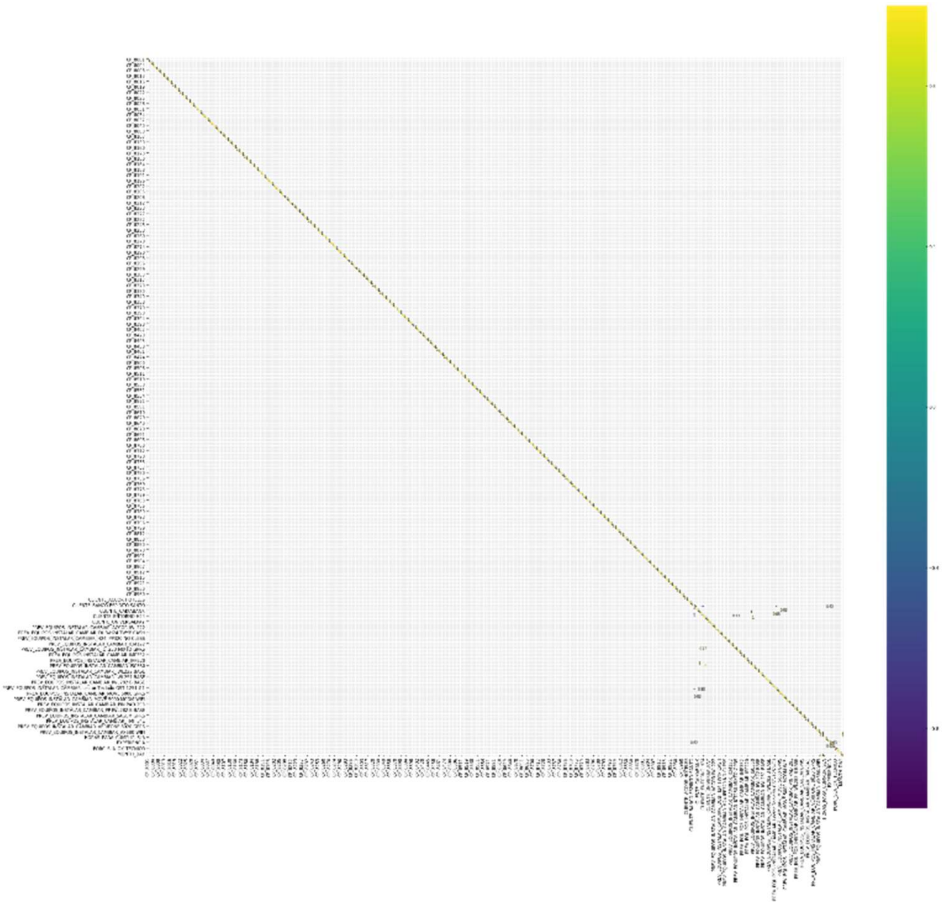


Ilustración 18. Mapa de correlación entre variables

3. Solución Tecnológica

3.1. Arquitectura general de la solución tecnológica

El objetivo de la solución SonaR diseñada para INDEA, es lograr la asignación óptima de recursos humanos, que en este caso son los técnicos de campo, a las incidencias que llegan del grupo Ingénico, mediante el uso de modelos de machine learning,

En el presente apartado, se presenta la solución piloto desarrollada para INDEA, la cual está orientada a facilitar la tarea de los coordinadores a la hora de asignar recursos para la resolución de las incidencias abiertas por el grupo Ingénico en la provincia de Barcelona. La solución se plantea como una aplicación integrada con el aplicativo Pez de INDEA, bajo el siguiente esquema.

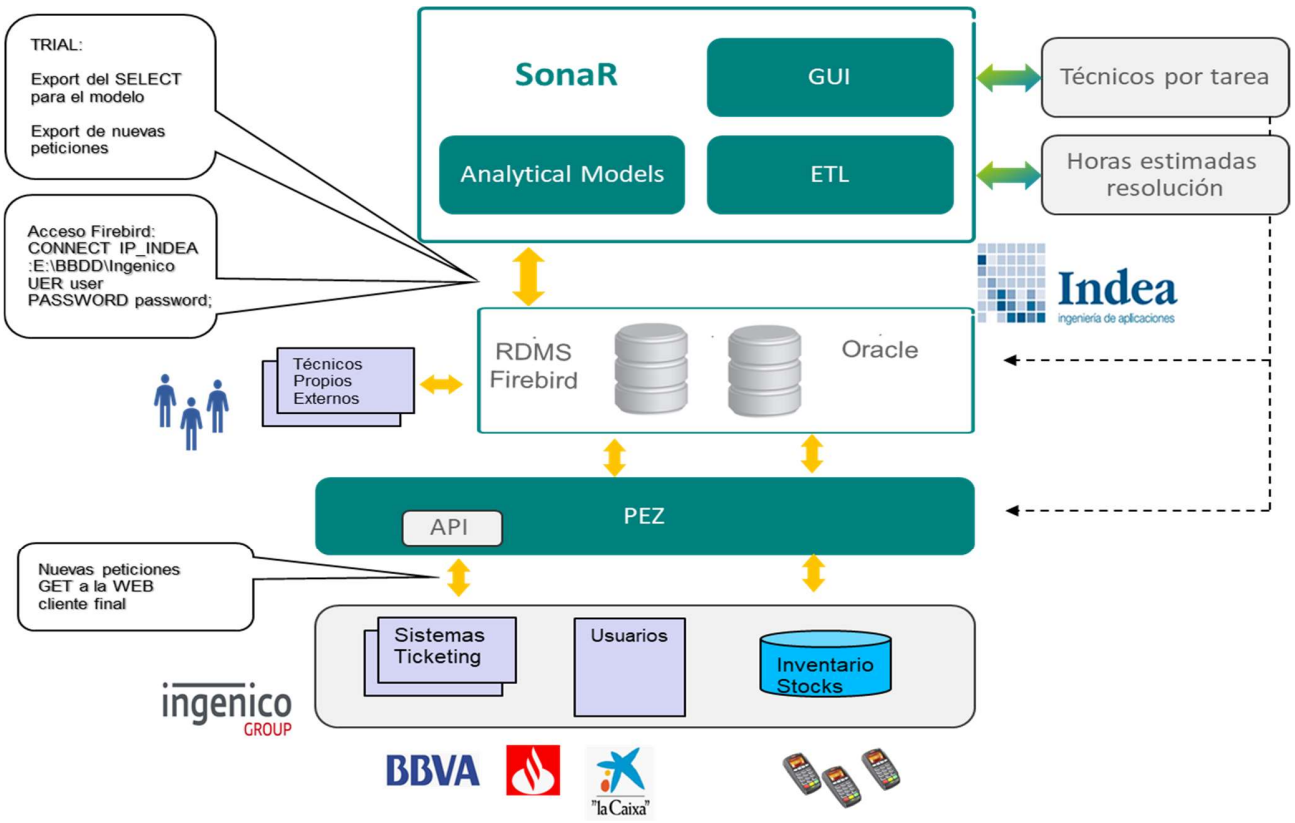


Ilustración 19. Esquema general de la solución SonaR

Ante la llegada de una nueva incidencia, el aplicativo Pez realiza una llamada al aplicativo SonaR con los datos identificativos de la incidencia. SonaR calcula en tiempo real para cada técnico el número de horas que necesitará para resolver la incidencia a partir de un modelo predictivo, y devuelve a Pez como respuesta un array con la recomendación de técnicos que previsiblemente tardarán un menor número de horas o estén mejor posicionados para la resolución de la incidencia.

En una primera fase, la información arrojada por la solución SonaR será usada por los coordinadores de INDEA para realizar la asignación de los técnicos a las incidencias, si bien a futuro y en función de los resultados obtenidos del periodo de evaluación del modelo, se podrá confiar en una asignación automática directamente mediante SonaR, reduciéndose de esta forma la carga de trabajo de los coordinadores y su aportación se limitará principalmente a incorporar novedades o decisiones estratégicas.

Para poder realizar la predicción, es preciso entrenar el modelo SonaR a partir de los datos históricos disponibles. Como se ha mencionado anteriormente, se dispone de un dataset con información detallada de las incidencias registradas el último año en la provincia de Barcelona para el grupo Ingénico, exceptuando las incidencias del grupo Ingénico para su cliente La Caixa.

El dataset, incluye, entre otros, datos geográficos, información del técnico que resolvió la incidencia, plazos y tiempos de resolución. Por tanto, a partir de dicha información histórica de cómo han efectuado y cómo ha funcionado el proceso de asignaciones de incidencias a técnicos en INDEA, se procede a entrenar el modelo de machine learning, mediante diferentes enfoques, los cuales se encuentran detallados en el siguiente apartado.

3.2. Modelo predictivo

En el presente apartado, se explican los diferentes enfoques de modelos predictivos analizados durante el desarrollo de la solución final desarrollada por SonaR.

3.2.1. Modelo predictivo basado en cumplimiento del SLA

En primera instancia, se plantea un modelo que pueda predecir si los técnicos cumplirán el SLA o no ante la llegada de una nueva orden de trabajo.

Entrenamiento del modelo

Como resultado del proceso ETL y analítica de las variables, se dispone de un dataset adecuado para proceder con el entrenamiento del modelo. Este dataset se procesa adecuadamente de tal modo que se obtienen:

- Datos de entrenamiento: El dataset utilizado para el entrenamiento, está compuesto de 383 columnas, distribuidas del siguiente modo:
    - 3 variables categóricas de tipo objeto que se utilizaran para entrenar diferentes enfoques de modelo:
        - SLA_OK: Variable que indica el cumplimiento o no cumplimiento de SLA.
        - SLA_TEC_GROUP: Variable que indica el grupo al cual pertenece el técnico que resolvió la incidencia de acuerdo a su porcentaje global de cumplimiento de SLA.
        - EXPERIENCIA_TEC_GROUP: Variable que indica el grupo al cual pertenece el técnico que resolvió la incidencia de acuerdo a su experiencia global.
    - 294 variables relacionadas con el Código Postal de la incidencia (linealizadas, manteniendo sólo las que aportan valor discriminatorio al modelo)
    - 14 relacionadas con clientes (linealizadas, manteniendo sólo las que aportan valor discriminatorio al modelo)
    - 59 relacionadas con los equipos (linealizadas, manteniendo solo las que aportan valor discriminatorio al modelo)
    - Fecha de entrada de la incidencia
    - Fecha tope de resolución de la incidencia
    - Tipo de Evento
    - Día de la semana en fecha de entrada
    - Día del mes en fecha de entrada

  Es importante destacar, que el dataset utilizado para el entrenamiento de este modelo, cuenta únicamente con variables de las que se dispondrá en el momento en el que se abra una nueva incidencia y el sistema Pez haga la llamada al modelo para realizar la asignación del técnico a la incidencia.
- Variable a predecir: La variable a predecir en este enfoque del modelo es el **cumplimiento del SLA**, para ello se entrenará el modelo con el conjunto de datos y se predecirá si las tareas asignadas cumplirán el SLA o no, usando como variable de entrenamiento SLA_OK.

Para entrenar el modelo, se divide el conjunto de datos en un conjunto de entrenamiento y otro de prueba, adicionalmente, el equipo de SonaR ha realizado entrenamientos con múltiples algoritmos con

el objeto de obtener el modelo con mayor tasa de acierto en la predicción al usar el cumplimiento de SLA como variable de entrenamiento. Los algoritmos utilizados han sido Random Forest (Regressor y Classifier), KNN (vecinos más cercanos) y Árboles de decisión.

Evaluación y Conclusiones

A continuación, se muestran los resultados obtenidos al utilizar los diferentes algoritmos para la creación del modelo basado en cumplimiento de SLA:

- Random Forest Regressor: Se utiliza este algoritmo con una profundidad de 1000 árboles de decisión obteniéndose scores del 19,39%
- Random Forest Classifier: Se utiliza este algoritmo con una profundidad de 150 árboles de decisión obteniéndose scores del 90,90%
- KNN (vecinos más cercanos): Se utiliza este algoritmo para un total de 10 vecinos, obteniéndose valores de score muy pobres, del orden del 13,55%
- Árboles de decisión: En este caso, se aplican diferentes parámetros al calcular el modelo, observándose que el mayor score se obtiene al aplicar GINI (85,79% de score con el dataset de prueba). A continuación, se muestra un resumen de los resultados obtenidos:

	Entropia(4)	Entropia(40)	Entropia	GINI(15)	GINI
Score (VC)	0.890351	0.897271	0.870565	0.895906	0.857895
Score (entrenamiento)	0.892351	0.903003	0.998747	0.898910	0.998747
Tiempo	0.227533	0.484302	1.305988	0.843551	2.374437

El modelo ofrece buenos resultados al utilizar Random Forest Classifier o árboles de decisión GINI, con valores de score superiores al 85%. Sin embargo, al hacer un análisis de la matriz de confusión generada en el caso particular del modelo basado en Random Forest Classifier, se puede apreciar una precisión muy alta en el caso de los SLA_OK, lo cual es el resultado de usar un dataset desbalanceado para el entrenamiento del modelo:

ANALISIS DE TEST

	precision	recall	f1-score	support
0	0.74	0.32	0.44	1278
1	0.91	0.98	0.95	8982
micro avg	0.90	0.90	0.90	10260
macro avg	0.82	0.65	0.69	10260
weighted avg	0.89	0.90	0.88	10260

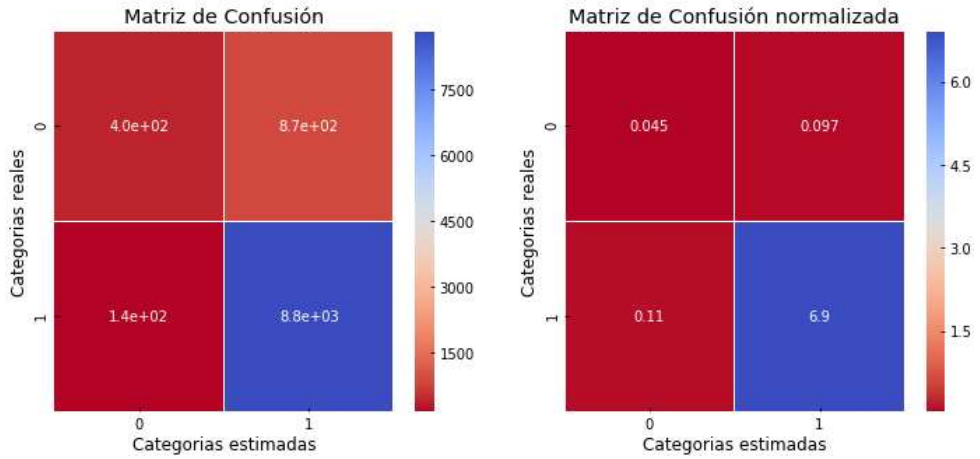


Ilustración 20. Matriz de confusión del modelo basado en Random Forest Classifier y SLA_OK como variable de entrenamiento

Este modelo debe entenderse como un apoyo a la asignación de las tareas, ya que únicamente ayuda a ver si los técnicos cumplirían o no el SLA, y no ayuda a la asignación automática de técnicos. Como consecuencia de ello, se plantea y desarrollan otros modelos a continuación.

3.2.2. Modelo Predictivo basado en Kmeans para caso de aprendizaje No Supervisado

Entrenamiento del modelo

Partiendo del mismo dataset de origen usado en el modelo anterior, se tiene lo siguiente:

- Datos de entrenamiento: El dataset utilizado para el entrenamiento, está compuesto de 383 columnas, y es equivalente al utilizado en el Modelo de Predicción del cumplimiento del SLA (véase apartado anterior para detalle de distribución de campos).

- Variable a predecir: La variable a predecir por el modelo es el cluster al que pertenecería cada incidencia de acuerdo al número de clusters establecidos.

Siguiendo la misma metodología que en el caso anterior, previo al entrenamiento, se dividen los datos disponibles en entrenamiento y test, para aplicar posteriormente la validación cruzada en la evaluación del modelo.

Evaluación y Conclusiones

En primer lugar, se utiliza el método del codo para conocer la cantidad de clusters óptimo que debe usarse en el entrenamiento del modelo, obteniéndose como se muestra en la Figura 3, que el número óptimo de clusters a utilizar es 11.

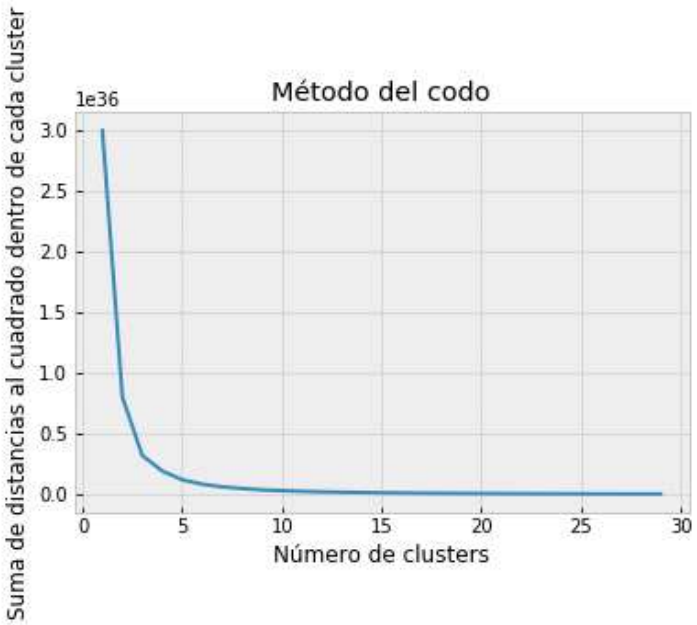


Ilustración 21. Método del codo aplicado a la obtención del número óptimo de clusters para el modelo de aprendizaje no supervisado basado en Kmeans

Posteriormente, se procede a entrenar el modelo con este número de clusters (11), ofreciendo buenos resultados obteniéndose un coeficiente de Silhouette de 0,543. Si bien se trata de un modelo de aprendizaje no supervisado, el número de clusters ideal en este caso es igual al número de rangos de experiencia diferentes que se definieron en la variable experiencia para el caso de los algoritmos de aprendizaje supervisado.

Sin embargo, dado que el objetivo del modelo es ayudar a los coordinadores a hacer una asignación óptima de recursos, se decide descartar este modelo como solución al problema descrito en el proyecto, ya que resulta imposible, al utilizar esta aproximación, conocer a qué cluster pertenecería cada uno de los técnicos, imposibilitando por tanto la asignación de los técnicos a las incidencias.

3.2.3. Modelo Predictivo basado en la experiencia de los técnicos

Entrenamiento del modelo

Partiendo del mismo dataset de origen usado en el modelo anterior, se tiene lo siguiente:

- Datos de entrenamiento: El dataset utilizado para el entrenamiento es equivalente al usado en los dos modelos anteriores (véase apartado 4.2.1 para detalle de los campos)
- Variable a predecir: La variable a predecir por el modelo es el **rango de la experiencia del técnico**, para ello se crean diferentes rangos a partir de la multiplicación de la variable EXPERIENCIA2 explicada en el apartado *3.6 Creación de Nuevos Campos* y la variable SLA_OK, resultando en la siguiente división del dataset de acuerdo al rango de experiencia del técnico (representando 10 el rango de técnicos de mayor experiencia, 1 los de menor experiencia y 0 aquellos donde no se cumple el SLA):

Siguiendo la misma metodología que en el caso anterior, previo al entrenamiento, se dividen los datos disponibles en entrenamiento y test, para aplicar posteriormente la validación cruzada en la evaluación del modelo.

En el presente modelo, el equipo de SonaR ha realizado entrenamientos con múltiples algoritmos con el objeto de obtener el modelo con mayor tasa de acierto en la predicción al usar la experiencia de los técnicos como variable de clasificación, entre ellos se encuentran los siguientes: Random Forest (Regressor y Classifier), KNN (vecinos más cercanos) y Árboles de decisión.

EXPERIENCIA_TEC_GROUP	CANTIDAD DE RESGISTROS
0	4171
1	322
10	9177
3	1918
5	5635

6	2779
7	4638
8	3465
9	2094

Evaluación y Conclusiones

A continuación, se muestran los resultados obtenidos al utilizar los diferentes algoritmos para la creación del modelo basado en la experiencia de los técnicos:

- Random Forest Regressor: Se utiliza este algoritmo con una profundidad de 1000 árboles de decisión obteniéndose scores del **32,18%**
- Random Forest Classifier: Se utiliza este algoritmo con una profundidad de 150 árboles de decisión obteniéndose scores del **53,55%**
- KNN (vecinos más cercanos): Se utiliza este algoritmo para un total de 10 vecinos, obteniéndose valores de score muy pobres, del orden del **10,54%**
- Árboles de decisión: En este caso, se aplican diferentes parámetros al calcular el modelo, observándose que el mayor score se obtiene al aplicar GINI (**50,06%** de score con el dataset de prueba). A continuación, se muestra un resumen de los resultados obtenidos:

	Entropia(4)	Entropia(40)	Entropia	GINI(15)	GINI
Score (VC)	0.293372	0.346491	0.477973	0.310624	0.500682
Score (entrenamiento)	0.288609	0.350223	0.997201	0.317641	0.997201
Tiempo	0.247397	0.861054	1.931471	1.282828	2.971735

Al realizar la comparativa de los resultados obtenidos para cada uno de los modelos entrenados con los diferentes algoritmos, se concluye que al utilizar como variable de clasificación la experiencia de los técnicos, el modelo basado en K-vecinos ofrece un resultado muy pobre, ofreciendo un porcentaje de acierto del 10,54%, posiblemente porque la cantidad de similitudes entre los diferentes registros no es suficiente para predecir los valores de experiencia, por lo que se decide descartar este modelo. El modelo basado en árboles de decisión ofrece mejores valores al utilizar GINI, obteniéndose un score de 50,06%, lo cual representa una mejora significativa respecto al modelo anterior. Sin embargo, el modelo basado en Random Forest Classifier, es el modelo que ofrece mejores valores de score (53,55%) para este caso de aprendizaje supervisado basado en la experiencia de los técnicos como variable de entrenamiento.

Es importante destacar, que, aunque los valores de acierto sean sólo del 47,55% y no del 98% (como es esperado para el cumplimiento de SLA), el modelo aporta valor al trabajo de coordinadores al

constituir una herramienta previa de ayuda para realizar una primera asignación de técnicos, la cual el coordinador podrá usar de base para realizar las asignaciones, minimizando el esfuerzo de los coordinadores en la realización de estas tareas.

3.2.4. Modelo Predictivo basado en tramos del tiempo de resolución

Los modelos presentados previamente sirven de soporte a la hora de determinar información complementaria de gran utilidad para el trabajo de los coordinadores. No obstante, se presenta un último modelo que pretende calcular el tiempo de resolución (en franjas o tramos de horas) que tardaría cada técnico, con el objetivo final de elaborar el ranking de técnicos que tardarían menos en resolver una orden de trabajo y por tanto actuarían a modo de sistema recomendador para el trabajo de los coordinadores, en línea con las necesidades indicadas por la Dirección de Indea al equipo de SonaR.

Entrenamiento del modelo

Para el entrenamiento de este modelo, se utiliza un dataset diferente al usado en los casos anteriores, compuesto por los datos que se indican a continuación:

- Datos de entrenamiento: conjunto de variables tras el proceso de limpieza. Se trata de 80 columnas, todas ellas de tipos enteros o float, distribuidas del siguiente modo:
    - 9 relacionadas con clientes (linealizadas, manteniendo solo las que aportan valor discriminatorio al modelo)
    - 33 relacionadas con los equipos (linealizadas, manteniendo solo las que aportan valor discriminatorio al modelo)
    - 30 relacionadas con los técnicos (linealizadas)
    - Tipo de evento
    - Distancia (calculado)
    - Número de incidencias abiertas
    - Experiencia del técnico en la resolución de dicha incidencia (calculado)
    - Día de la semana, día del mes y mes

  Destacar que en el dataset se eliminan variables que aportarían gran información al modelo, pero de las cuales no se dispondrá a la hora de utilizar el modelo en tiempo de ejecución,

tales como Horas para cumplir SLA, fechas de resolución, incidenciado, tiempo de resolución, etc.

- Variable a predecir: tiempo de resolución de la incidencia. Esta variable se utilizará para crear rangos de tiempo en los cuales se intentará categorizar a los técnicos.

Previo al entrenamiento, se dividen los datos disponibles en entrenamiento y test, para aplicar posteriormente la validación cruzada en la evaluación del modelo. Una vez llevada a cabo esta separación, el equipo de SonaR ha realizado el entrenamiento basándose en primera instancia en árboles de decisión, concluyéndose la necesidad de utilizar redes neuronales en una segunda fase. El proceso se detalla a continuación.

Evaluación y conclusiones

Al entrenar el modelo con árboles de decisión para predecir el número exacto de horas que invierte cada técnico en la resolución de una incidencia, los resultados obtenidos son deficientes. Por otra parte, Indea requiere para que el modelo sea de utilidad que el modelo desarrollado permita dar predicciones sobre en qué rango de horas la incidencia puede estar resuelta por cada uno de los técnicos, no necesariamente el número específico de horas. Este hecho ha supuesto un reto para el equipo, pues la variable objetivo del modelo era continua (el número de horas en que cada incidencia se cerró), debiendo discretizar esta variable en diferentes rangos de horas. Considerando lo anterior y el hecho de que no es necesario conocer con total precisión los tiempos de resolución de cada incidencia para conseguir los objetivos del proyecto, se ha llevado a cabo un proceso iterativo de categorización de tiempos y pruebas, con el objeto de mejorar los resultados.

En los siguientes apartados, se describen los pasos seguidos y los resultados obtenidos.

a) Predicción de franjas horarias

En primer lugar, se ha planteado la agrupación en franjas o tramos de horas. A modo de ejemplo: si se utilizan tramos de 6 horas, en 2 días, las resoluciones se dividirán en 8 tramos: tramo 1 (de 0 6 horas), tramo 2 (de 7 a 12 horas), tramo 3 (de 13 a 18 horas), y así sucesivamente hasta el tramo 8 (de 42 a 48 horas). Si el sistema es capaz de predecir en qué tramo horario resolvería cada técnico una nueva incidencia, aportaría información muy valiosa al coordinador de cara a la asignación del técnico óptimo. Si en el mejor tramo horario coinciden varios técnicos, SonaR propondrá a todos, siendo el coordinador el responsable de decidir el más adecuado a partir de su conocimiento de negocio y el contexto particular.

El objetivo del modelo pasa por tanto a ser la predicción del tramo de horas. Lo que se está realizando con esta acción es, en cierto modo, "categorizar" o discretizar el resultado deseado, es decir, en lugar de contar con un alto número de valores distintos de horas a predecir, se cuenta con un número limitado de tramos de horas.

Lógicamente, no se conoce a priori cuál es la duración ideal de tramo de hora. Llevando el concepto del tramo al límite de su granularidad, se obtendrían tramos de 1 hora, por lo que se volvería a tener la misma variable "continua" original. Un hecho importante a destacar es que, según se hacen más grandes los tramos (mayor cantidad de horas por tramo), se obtienen estimaciones más "gruesas", que aportan por tanto una información menos valiosa al coordinador.

Para calcular la duración del tramo de hora a aplicar, se crea un bucle que se va iterando, transformando la variable *Tiempo de Resolución* en distintos números de tramos de diferente duración. Se trata por tanto de obtener un compromiso entre el tamaño de la franja de horas y el rendimiento de los modelos.

Dentro de cada iteración, se realiza a su vez el entrenamiento con múltiples árboles, variando el tipo (GINI o entropía) y el número de nodos, usando para ello nuevamente iteración y simulaciones con múltiples ocurrencias.

En general, se observa en los resultados que, a partir de un cierto número elevado de hojas provoca *overfitting* y empeora los resultados del rendimiento del modelo, obteniéndose habitualmente los mejores resultados en árboles con un número de hojas entre 100 y 300.

Los resultados obtenidos son, en una primera aproximación, lógicos, dado que, según se aumenta el número de tramos de horas (cada uno de ellos, de menor número de horas), se obtienen resultados (*scoring*) de los modelos inferiores (a los modelos les cuesta más precisar tramos de horas más pequeños). De este modo, en la iteración de los bucles, se va obteniendo información que manifiesta el siguiente compromiso:

- Si se usan tramos de horas muy grandes, la predicción no aporta valor. Por ejemplo, si se usan escalones de 24 horas, las predicciones apenas tendrían validez para INDEA, ya que todos los técnicos suelen resolver las incidencias en 1 ó 2 días y los valores predichos por el modelo corresponden a los primeros tramos (de 0 a 24 horas, de 25 a 48 horas, o de 48 a 72) en la mayoría de los casos. Por tanto, con tramos de mucha duración prácticamente todas las ocurrencias del dataset de entrenamiento caen en las mismas franjas, y el modelo se comportará de este modo al predecir valores (situará a todos los técnicos en el primer tramo de incidencias de tipo mantenimiento y en el segundo para nuevas instalaciones).

Se muestra a continuación la distribución de las ocurrencias en los datos de entrenamiento utilizando tramos de 24 horas durante 1 semana, para la cual se obtiene un score muy elevado, del 87,8% (usando el método GINI y 150 hojas de profundidad de árbol):

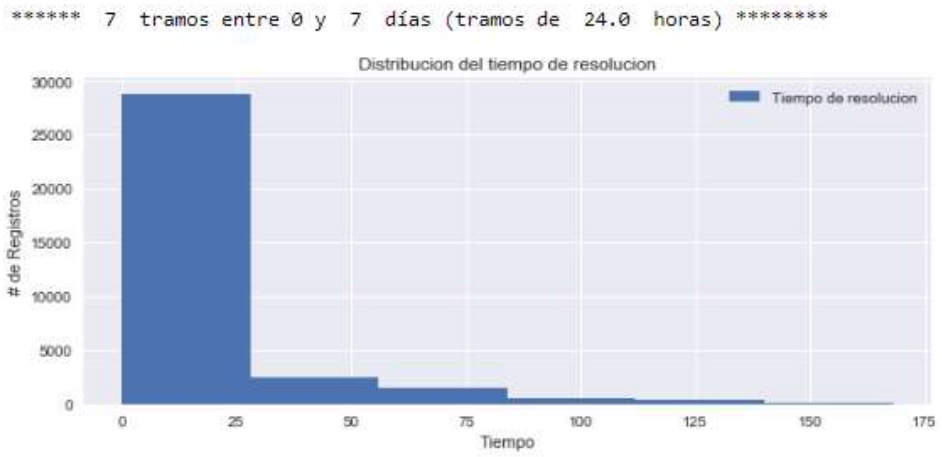


Ilustración 22. Distribución de las ocurrencias en los datos de entrenamiento utilizando tramos de 24 horas durante 1 semana

- Por el contrario, si se usan tramos de horas muy pequeños, los modelos empiezan a disminuir su *scoring*. Por ejemplo, con tramos inferiores a 9 horas el sistema aportaría valor a INDEA . Sin embargo, el *scoring* de ese modelo (GINI, 290 hojas) es del 59%, lo cual reduce mucho la fiabilidad del sistema.

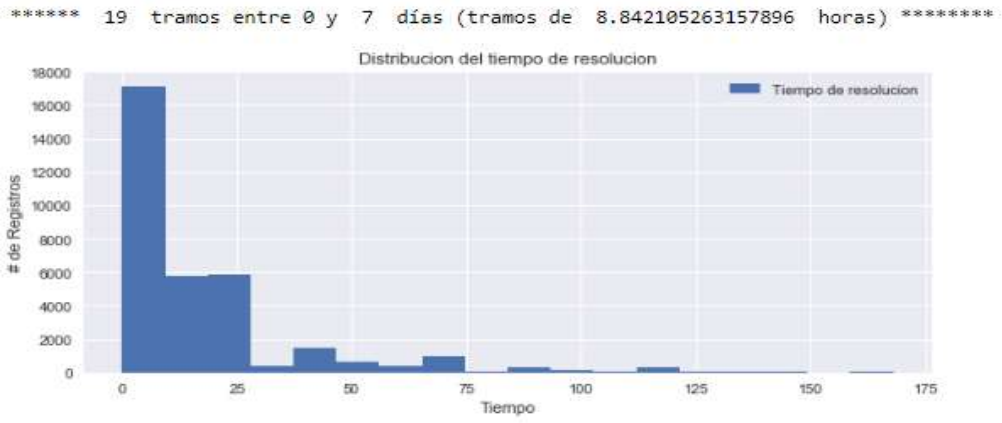


Ilustración 23. Distribución de las ocurrencias en los datos de entrenamiento utilizando tramos de 9 horas durante 1 semana

A continuación, se presenta un resumen de algunos resultados obtenidos tras sucesivas iteraciones y pruebas (el equipo de SonaR dispone de un sistema parametrizado con el que ha efectuado múltiples pruebas):

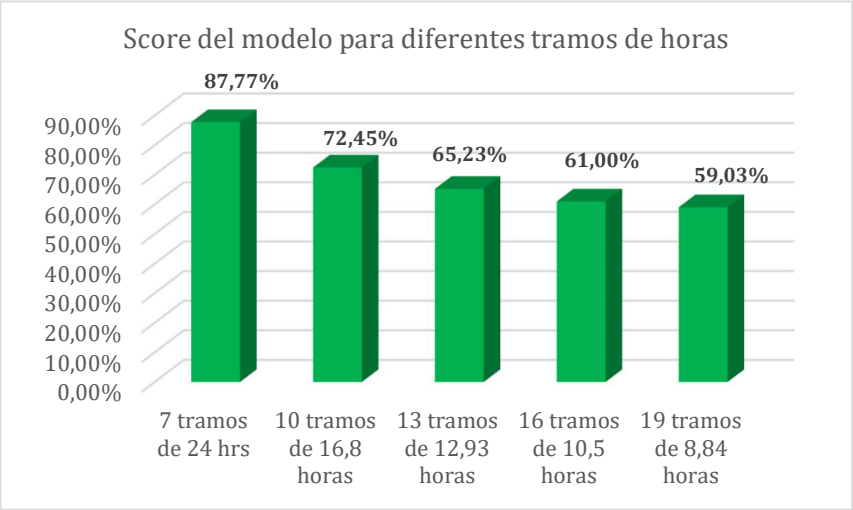


Ilustración 24. Representación gráfica del score del modelo basado en árboles de decisión y variable de entrenamiento tiempo de resolución para diferentes tramos de horas

De este modo, la utilización de tramos de 10.5 horas sería adecuada (16 tramos en una semana), ya que aportaría información suficiente al coordinador para su toma de decisiones, y el *score* del modelo es aceptable (61%).

Finalmente, se crea un conjunto de tramos horarios con "sentido para el negocio":

- Tramo 1: de 0 a 2 horas
- Tramo 2: de 2 a 18 horas
- Tramo 3: de 18 a 24 horas
- Tramo 4: de 24 a 48 horas
- Tramo 5: de 48 a 100 horas
- Tramo 6: más de 100 horas

Los límites de 24 y 48 horas son claves en el modelo pues suponen los límites para considerar a una petición como resuelta o no según el tipo de petición. Así, como se ha mencionado en apartados anteriores, si se trata de una petición de instalación de equipo, existen 48 horas para que sea resuelta, reduciéndose este tiempo a 24 horas en el caso de tratarse de una petición de mantenimiento. Se calcula el árbol que maximiza el rendimiento con estos tramos, y se obtiene un árbol GINI con 290

hojas, que constituye el mejor modelo detectado para este caso, obteniendo un score del 49,82%. Se realizan simulaciones adicionales con este modelo de tramos y distintos algoritmos de aprendizaje; los mejores resultados se dan con un modelo *Random Forest Clasiffier*, que con 500 árboles obtiene un score de 53,58%.

b) Explicabilidad de las variables

A diferencia de otros métodos de aprendizaje supervisado, una de las grandes ventajas de los árboles de decisión es que se puede consultar cómo están operando internamente.

Para realizar el análisis de la relevancia de las características, se utiliza una función basada en el error cuadrático medio del árbol mejor posicionado. Los resultados que obtienen al aplicar este análisis a los modelos de árbol de decisión y *Random Forest Classifier* son esclarecedores: las variables más relevantes a la hora de determinar el tramo horario en el que se sitúa la resolución de una incidencia son el tipo de evento (si es mantenimiento o nueva instalación), el día de la semana, el mes, el código postal, o el equipo a instalar, entre otras, dándose ligeras diferencias entre ambos modelos. A continuación, se presenta una muestra de los valores numéricos obtenidos en relación a la representatividad de las características para ambos modelos:

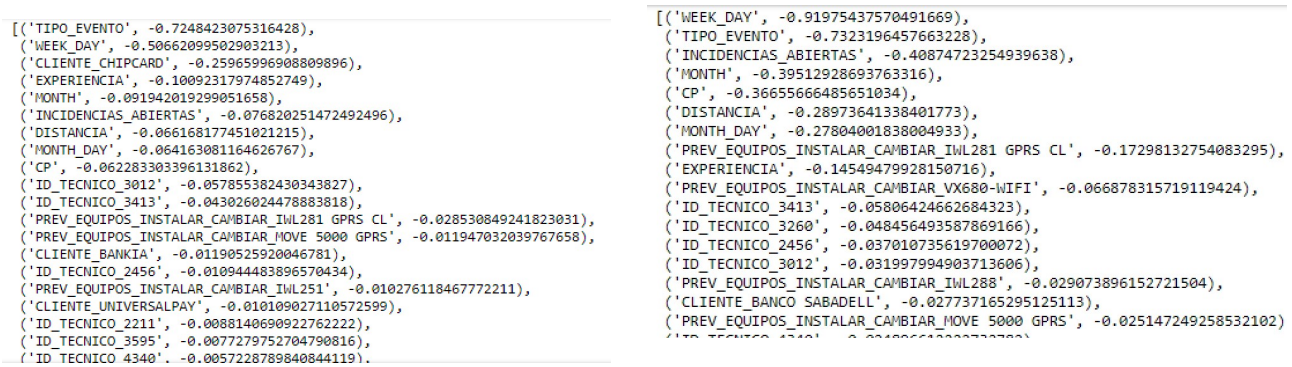


Ilustración 25. Relevancia de las variables en la decisión de asignación de franja horaria con el mejor árbol (GINI de 290 hojas), a la izquierda, y con Random Forest, a la derecha

Estas variables han sido contrastadas con la empresa, pareciendo totalmente razonable su utilización y, mostrando de una manera analítica, la forma en que las peticiones son asignadas.

En el caso del modelo con árbol de decisión, las variables resultan de escasa utilidad de cara a discernir entre técnicos. En este caso, la analítica aplicada sobre los datos que almacena la empresa INDEA arroja como conclusión que las variables más determinantes a la hora de determinar la franja de tiempo

en que se solucionará una incidencia son comunes a todos los técnicos. Por tanto, prácticamente todos los técnicos se situarán en la misma franja horaria de resolución de incidencias, y el modelo no discernirá de forma eficiente entre unos y otros; SonaR podrá informar a Pez del tiempo previsto de resolución (franja horaria) en función de la incidencia recibida, y en la mayor parte de los casos recomendará un conjunto amplio con la mayoría de técnicos, que estarán posicionados en la misma franja horaria prevista, no técnicos concretos idóneos para llevarla a cabo por precisar menor tiempo de resolución que los demás. Esto se ha podido comprobar realizando simulaciones de llamadas a SonaR por parte de PEZ con datos variables, y calculando la franja predicha para cada técnico, obteniendo para el modelo del árbol la misma franja para todos los técnicos en todas las pruebas efectuadas.

Si bien en gran medida estos resultados son comunes al modelo Random Forest, es cierto que en este modelo sí considera la influencia de la situación particular de cada técnico. De este modo, se aprecia que el número de incidencias abiertas es la tercera variable más relevante en su proceso de predicción. Este efecto por el cual este modelo sí es capaz de discernir ligeramente entre técnicos se puede comprobar realizando la misma simulación que la efectuada anteriormente con el modelo de árbol de decisión. A continuación se presentan los resultados de simular la misma llamada desde Pez, observándose que en este caso el modelo sí que aprecia que dos técnicos tardarían menos en resolver la incidencia, situándolos en un tramo inferior al resto (técnicos 2623 y 3256):

```
* Predición técnico  2211  con Random Forest:  3
* Predición técnico  2349  con Random Forest:  3
* Predición técnico  2392  con Random Forest:  3
* Predición técnico  2456  con Random Forest:  3
* Predición técnico  2488  con Random Forest:  3
* Predición técnico  2623  con Random Forest:  2
* Predición técnico  3012  con Random Forest:  3
* Predición técnico  3174  con Random Forest:  3
* Predición técnico  3182  con Random Forest:  3
* Predición técnico  3189  con Random Forest:  3
* Predición técnico  3256  con Random Forest:  2
* Predición técnico  3260  con Random Forest:  3
* Predición técnico  4336  con Random Forest:  3
```

Ilustración 26. Resultados de simular llamada de Pez a SonaR (modelo RF)

c) Simulación reduciendo las variables comunes

Con el objeto de mejorar la discriminación entre los técnicos, se realizaron nuevos entrenamientos de árboles para distintas franjas horarias eliminando las variables comunes que resultaban ser más determinantes a la hora de discernir las franjas horarias:

- En primer lugar, se elimina el efecto de la variable más determinante (TIPO_EVENTO) dividiendo el dataset en dos, datos correspondientes a incidencias/mantenimientos (SLA 24 horas) y datos correspondientes a nuevas instalaciones (SLA 48 horas). Al iterar en los tramos, se observa de forma evidente que la distribución de los casos de entrenamiento responde a este patrón, situándose siempre el conjunto más grande de ocurrencias cercano a esos dos momentos, dependiendo del tramo, y siendo muy diferente en el caso del entrenamiento de un dataset o del siguiente. No obstante, al realizar este ejercicio se obtienen conclusiones similares a las expuestas en el apartado anterior: las variables más determinantes continúan siendo comunes a todos los técnicos (día de la semana, mes del año, o tipo de equipo, entre otros), y los modelos no sirven para discriminar de forma clara entre técnicos, obteniéndose una ligera discriminación nuevamente con Random Forest. Se concluye que el sistema se está comportando de forma lógica, dado que ha descubierto de los datos que las incidencias se tardan en resolver más o menos en meses de verano, los viernes o para ciertos tipos de equipo, y en menor medido debido a la influencia de características propias de cada técnico. Esas variables causan que las incidencias se sitúen en uno u otro tramo horario de resolución, relegando a un segundo plano la influencia que puedan tener otras variables diferenciadoras de los técnicos (experiencia, distancia desde la que parte, incidencias que tiene abiertas en cada momento…).
- Con el objetivo de poder discernir entre los técnicos, se eliminan del dataset más variables comunes a todos ellos (según lo mencionado anteriormente, relacionadas con la fecha, el equipamiento, etc). Sin embargo, el dataset resultante queda mermado y el comportamiento de los árboles tras el entrenamiento no es el esperado, obteniéndose *scorings* en torno al 30%. Por tanto, se concluye, que los modelos de árboles no son capaces de predecir con fiabilidad la franja horaria en la que se resolverá una incidencia únicamente a partir de variables como la experiencia del técnico, la distancia a la que se encuentra y cuántas incidencias esté procesando en ese momento. Esto se observa incluso si se intentan predecir en el entrenamiento franjas horarias más "gruesas" (de 24 horas, por ejemplo).

d) Interpretación y análisis de la utilidad del modelo

El modelo consigue predecir correctamente el 53.58% de las franjas horarias establecidas. Este porcentaje puede parecer pequeño pero los resultados son considerados satisfactorios por diferentes motivos que se detallan a continuación.

En primer lugar, es muy difícil de predecir con exactitud el tiempo en horas en base a la poca información disponible. Existen muchas variables que determinan que una incidencia se cierre antes o después, siendo una de ellas también cuándo el propio instalador da por cerrada la incidencia,

pudiendo darse diferencias e inexactitudes en los datos almacenados entre la hora de cierre real y cuando éste la introduce en el sistema.

En segundo, si se analizan los resultados obtenidos al aplicar el modelo la predicción sobre el conjunto de test, se obtienen resultados de gran valor para Indea. Las siguientes gráfica y tabla muestran estas diferencias:

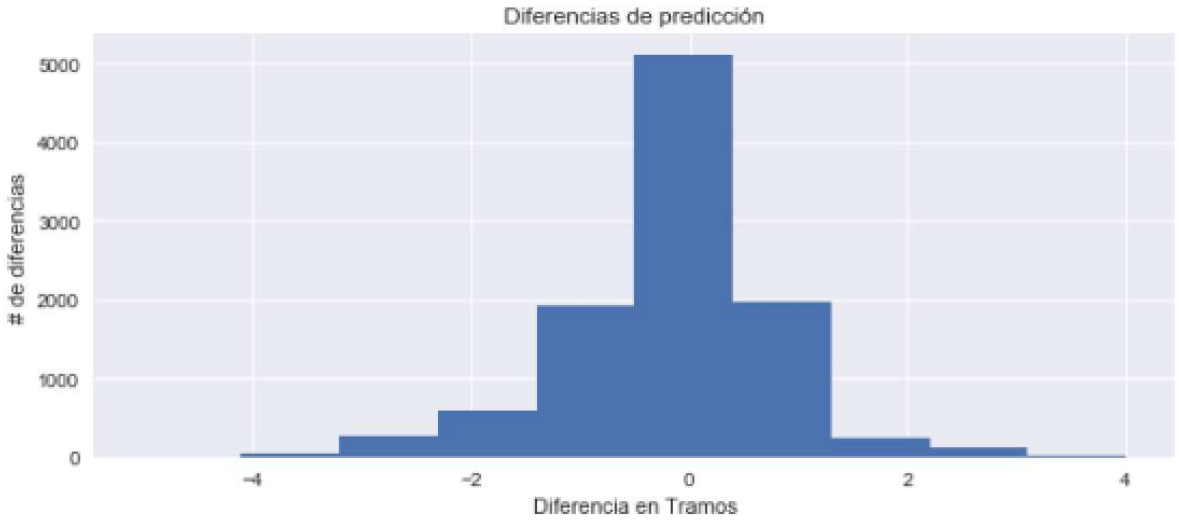


Ilustración 27. Diferencia entre la predicción del modelo y los datos de test (representación visual del histograma)

Diferencia Y pred - Y test	Número de casos	Porcentaje
0	5497	53,58%
-1	1846	17,99%
-2	512	4,99%
-3	284	2,77%
-4	52	0,51%
1	1713	16,70%
2	244	2,38%
3	92	0,90%
4	19	0,19%

5	1	0,01%

Tabla 4. Diferencia entre la predicción del modelo y los datos de test (datos cuantitivos)

Como se puede comprobar:

- De las ocasiones que el sistema no acierta el tramo (un 46,42%), se equivoca únicamente por un escalón en la mayor parte de los casos (34,69%). Los casos de error grande de cálculo de franja horaria (valores mayores de diferencias) son anecdóticos y pueden ser debidos a anomalías o comportamientos excepcionales.
- El sistema es "pesimista", en el sentido que predice en mayor número de ocasiones franjas de tiempo por debajo de las franjas reales. En un 26,26% (suma de las diferencias negativas) de los casos el modelo predice que el técnico tardaría menos tiempo de lo que tardó en realidad.

Un análisis alternativo / complementario se encuentra al analizar en detalle la matriz de confusión y el porcentaje de categorías acertadas con diferentes *notches*:

```
[[ 885 1213   57   42   17    1]
 [ 630 3133  305  103   49    2]
 [ 125  921  587  118   80    1]
 [  50  296  152  397   72    4]
 [  32  214   85  126  472    5]
 [   0   20   20    6   17   23]]
```

Ilustración 28. Matriz de confusión

notch 1	5497	53,58%
notch 2	9056	88,27%
notch 3	9812	95,63%
notch 4	10188	99,30%
notch 5	10259	99,99%
notch 6	10260	100,00%

Tabla 5. Porcentaje de categorías acertadas a diferentes notches de la matriz de confusión.

Como se puede apreciar, el modelo acertaría el 88,27% de las franjas horarias en 1 notch de la matriz de confusión en la muestra de test (es decir, considerando que acierta o se equivoca únicamente en un tramo). Este dato es equivalente al calculado anteriormente mediante el histograma de diferencia de los valores predichos. Es decir, el modelo o bien acierta, o bien se aproxima al tramo de tiempo de resolución de la orden de trabajo en un porcentaje por muy alto.

Adicionalmente, observando la matriz de confusión se puede obtener un dato relevante para garantizar la validez del modelo: el impacto negativo de las equivocaciones del modelo. SonaR únicamente

resulta perjudicial en los casos en que se equivoca prediciendo que un técnico tardaría menos en solucionar una incidencia que lo que realmente tarda. El modelo estaría haciendo perder dinero a la empresa por enviar a un técnico prediciendo cumplimiento de SLA y dándose finalmente incumplimiento del mismo; en el caso de las incidencias/mantenimientos, supone pasar de un tramo predicho entre el 1 y el 3 a un tramo real igual o superior al 4. Se ha realizado este cálculo para las incidencias (mantenimientos), resultando únicamente un 8.8% de los casos.

Otras evaluaciones y trabajos a futuro. Modelo de redes neuronales

Se han planteado simulaciones con redes neuronales con el objeto de mejorar los resultados, pretendiendo encontrar un modelo que no disminuya tanto su rendimiento cuando se reducen los tramos de duración lo suficiente como para que aflore la influencia de las variables específicas de cada técnico en mayor medida que como sucede en Random Forest. Si bien es cierto que las redes neuronales limitan en gran medida el conocimiento interno que se tiene del comportamiento del modelo, llegados a este punto de comprensión de los datos disponibles y el comportamiento de los árboles, el objetivo es la mejora de prestaciones del modelo aunque no sea tan explicativo.

Se realiza un proceso equivalente al mencionado anteriormente para comprobar si la aportación de las redes neuronales al problema descrito anteriormente es suficiente y permite mejorar los resultados de cara a la predicción de franjas horarias como trabajo de soporte a los coordinadores. Los primeros resultados obtenidos son prometedores; incluso con tramos de 6 horas, se obtienen accuracies muy elevadas (86,88%) prácticamente desde las primeras épocas.

Elección y entrenamiento del modelo

Para entrenar el modelo se hace uso de la librería Keras con TensorFlow como backend. Se prepara la secuencia de los modelos con varias capas rectificadoras (‘relu’, ‘tanh’) y una capa final que inicialmente usaba activación con función sigmoide tras haber definido un tiempo de umbral de resolución y categorizado la variable en función de dicho valor (‘1’, ‘0’).

Para el caso de clasificación usamos la métrica *accuracy*. Se entrenan los modelos con diferente número de capas (`largo de la red), diferentes entradas (ancho de la red) y con diferentes números de ciclos y tamaños del batch de variables

La clasificación binaria da buenos resultados, pero siendo dicho modelo de aplicación relativa (lo que explicaremos a continuación en la simulación del modelo), a pesar de su alto grado de *accuracy*, entrena la red neuronal usando la variable categórica asociada al tiempo de resolución. Se eligen para

ello dos modelos, uno con dos capas rectificadoras y una final sigmoide, y otro con dos capas rectificadoras y una final *softmax*.

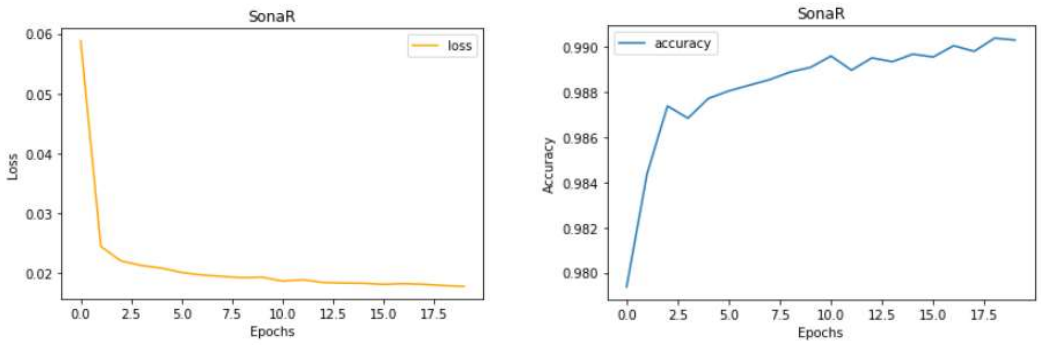


Ilustración 29. Evolución del accuracy y loss en modelo de redes neuronales

El modelo en este último caso necesita categorizar los tiempos de resolución, establecer el loss mediante categorical_crossentropy y establecer como métrica el categorical_accuracy. Los niveles de accuracy son altos y el loss va disminuyendo con el entrenamiento de la red, según se muestra en el siguiente esquema:

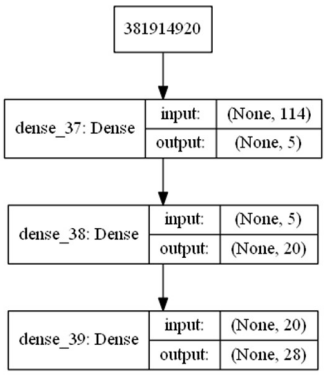


Ilustración 30. Modelo redes neuronales

Simulación de datos de entrada

Se observa que los casos donde falla la predicción coinciden con los tiempos más altos de la muestra, probablemente por la distribución no uniforme de tiempos de resolución en el conjunto de datos. Se prueba únicamente en los casos no incidenciados y el modelo funciona, sin embargo al no ser de aplicabilidad de momento en los casos incidenciados se propone dejar el modelo on-hold hasta que el modelo permita incluir ambos casos, y continuar con otra aproximación más adecuada.

Uso de red neuronal en próximas implementaciones

Este modelo con pocos cambios podría servir para evaluar un rango de variables (equipos, técnicos), lo que podría ser de aplicación futura para evaluación adicional de competencia de técnicos o para evaluación de dificultad de servicio para determinados equipos.

Sin embargo, en su estado actual, el modelo al efectuar las predicciones arroja valores poco definitivos, causa por la cual por el momento no se opta como solución final para el problema. Por tanto, la aplicación de redes neuronales es una de las líneas de trabajo claras en el proceso de expansión de SonaR a futuro al resto de clientes y geografía nacional.

3.3. Explotación del modelo

Partiendo de la idea de que el modelo seleccionado para dar solución al problema planteado en este proyecto es el modelo predictivo basado en tramos de tiempo de resolución, durante la fase de explotación del modelo, SonaR se ejecuta ante la recepción de nuevas incidencias siguiendo el siguiente procedimiento:

- Pez realiza una consulta a SonaR ante la llegada de una nueva incidencia, con los siguientes campos disponibles en ese momento: fecha de entrada, cliente y ubicación del mismo, equipo a instalar o cambiar, tipo de evento y datos de los técnicos (número de incidencias abiertas y ubicación de la que parten), todos ellos datos disponibles en Pez en el momento del alta de una incidencia.
- SonaR calcula en tiempo real los nuevos campos necesarios para la aplicación del modelo: experiencia de los técnicos (para el tipo de equipo incidentado y cliente concreto), día del mes, día de la semana y mes (a partir de la fecha) y distancia (a partir de la longitud y latitud de la anterior incidencia, medida en kilómetros), usando las mismas fórmulas indicadas en el proceso ETL según lo especificado en el apartado 2.6 del presente informe.

Cabe destacar que para la integración los sistemas Pez – SonaR es necesario que Indea desarrolle en Pez las llamadas a SonaR, tanto para enviar los datos ante la llegada de una nueva incidencia como

para interpretar la respuesta de SonaR (matriz de los técnicos propuestos que tardan menos tiempo en resolver la incidencia)

A fecha de presentación del presente Proyecto Fin de Máster esta integración aún no ha sido desarrollada, debido a la imposibilidad del equipo de INDEA de afrontar la integración por carga de trabajo, si bien existe la firme convicción del Director de Operaciones de apadrinar el proyecto y dedicar los recursos necesarios para efectuar la prueba piloto. Para la realización de las pruebas pertinentes, se ha realizado un *mock* o simulación de cómo sería el funcionamiento de la llamada de este servicio web.

3.4. Mantenimiento del modelo

Se estima que será necesario re-entrenar el modelo como mínimo mensualmente, dado que el volumen de datos recibido al mes es significativo y podría influir en el comportamiento del modelo. Por ejemplo, durante un mes la experiencia de un técnico en un cliente o con un nuevo equipo puede variar significativamente.

En el caso de que la asignación de los técnicos fuese completamente automática, es posible que esta realimentación no fuese necesaria, ya que el modelo se seguiría comportando del mismo modo con los nuevos datos decididos. Pero dado que se han planteado una solución a modo de soporte a la operación de los coordinadores, éstos irán realizando pequeñas correcciones a las propuestas del modelo, descartando en ciertas ocasiones los técnicos propuestos por SonaR, creando de este modo datos relevantes de cara al reentrenamiento del modelo, lo cual posibilitará ir superando progresivamente el porcentaje de éxito obtenido inicialmente.

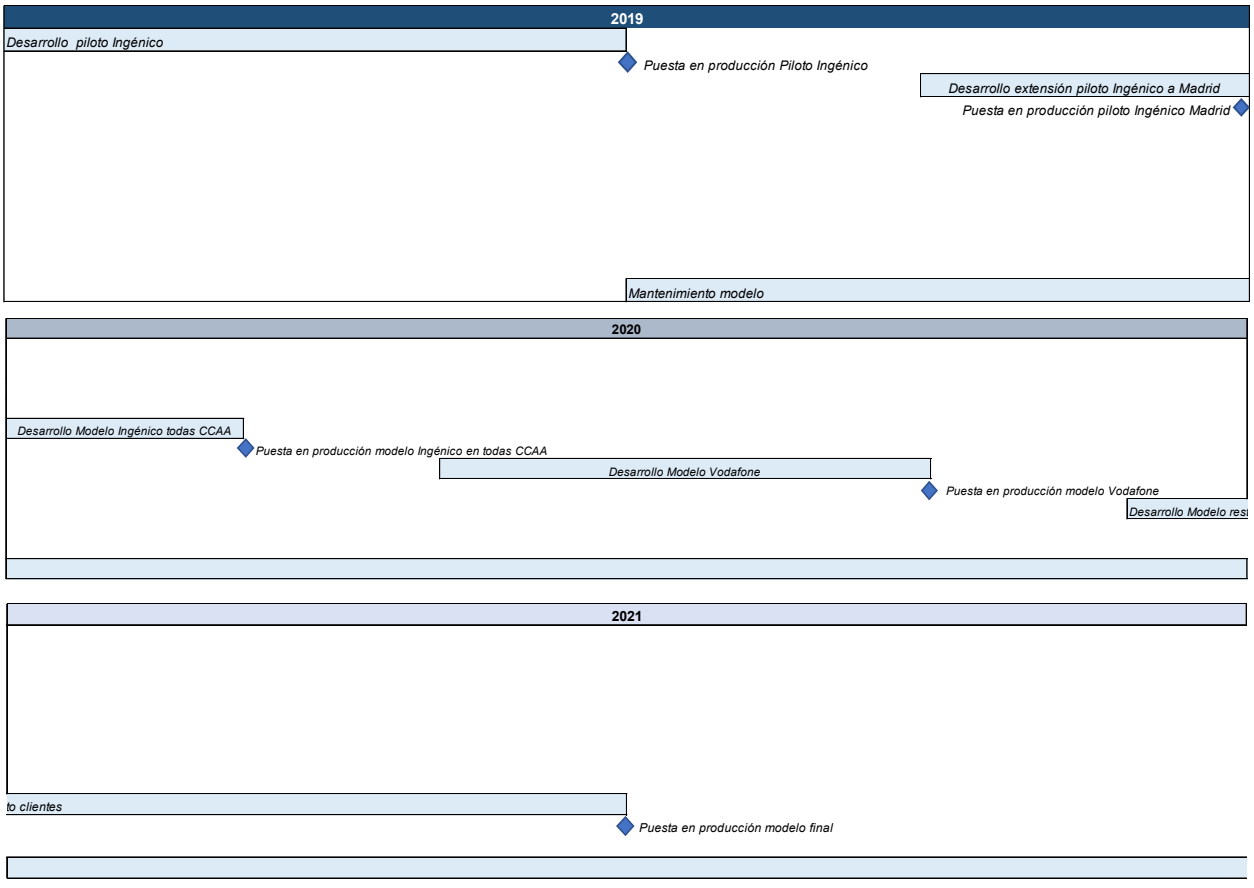
Adicionalmente, es necesario considerar el reentrenamiento mensual del modelo debido a la posible aparición de nuevos modelos de equipos, así como también el alta y/o baja de los diferentes técnicos en plantilla.

3.5. Expansión del modelo

El siguiente esquema refleja la expansión del modelo, con los principales hitos de ejecución previstos:

- 1º Semestre 2019: desarrollo piloto para el cliente Ingénico en Barcelona
- Inicio de mantenimiento del modelo. Una vez puesto en producción, se plantea un mantenimiento evolutivo y correctivo del modelo de forma continuada

- 2º Semestre 2019: desarrollo de la extensión del piloto para el cliente Ingénico a Madrid
- 1º Semestre 2020: desarrollo de la extensión del piloto para el cliente Ingénico a resto de Comunidades Autónomas
- A lo largo del año 2020: desarrollo de la extensión del piloto al cliente Vodafone
- Finales año 2020 y principios 2021: desarrollo de la extensión del piloto a resto de clientes de Indea



4. Análisis y proyección económica

4.1. Planteamiento general del *business case*

Se plantea el *business case* considerando únicamente el beneficio del sistema SonaR. Por tanto, sólo considera los ingresos y costes directamente relacionados con el sistema, y no otros generales, como pueden ser instalaciones, coste de personal de Indea (no directamente relacionado con el sistema), alquileres, etc. Según esta consideración, se han identificado los principales elementos a considerar en el análisis económico:

- Ingresos obtenidos por Indea como consecuencia de la implantación del modelo:
    - Ahorros por penalizaciones evitadas: El modelo tendrá como consecuencia un aumento de cumplimiento de SLA, que tiene como consecuencia directa un menor coste para Indea de las penalizaciones por incumplimientos del mismo.
    - Ingresos por aumento del número de incidencias gestionadas: Una mejor asignación de los técnicos posibilitará la reducción del tiempo para la resolución de las incidencias, y por tanto un aumento de la capacidad de gestionar un mayor número de incidencias con los mismos recursos en el mismo tiempo.
    - Ahorro en coste de personal: Según se vaya automatizando el proceso de asignación de incidencias a técnicos, los coordinadores se verán liberados de esa tarea, pudiendo dedicar su tiempo a actividades de mayor valor añadido. Se espera que inicialmente la herramienta actúe como soporte a la operación, y que con posterioridad vaya confiándose progresivamente en SonaR hasta un funcionamiento completamente autónomo o bajo la supervisión puntual de casos específicos.

- Costes de implantación del modelo
    - Costes equipo desarrollo: El principal coste del proyecto es el de los recursos humanos del equipo de SonaR.
    - Se asume que la infraestructura existente en Indea puede asumir sobradamente las necesidades de la implantación del sistema SonaR. No obstante, se reserva una partida menor a largo plazo para reforzar esta infraestructura, ante el posible aumento de necesidad de procesamiento y almacenamiento según se escale el proyecto.

4.2. Hipótesis

Las principales hipótesis tenidas para el cálculo del *business case* son las siguientes:

4.2.1. Hipótesis en relación a los ingresos

- **Aumento de cumplimiento de SLA**

  Se estima el aumento del porcentaje de cumplimiento de SLA a partir del aumento que se da en los meses de verano, calculable en base al dataset disponible.

  Según los datos con los que se cuentan, en el periodo de un año el cumplimiento medio del SLA es del 87.80373695137285%. Si se recalcula este porcentaje excluyendo el mes de Agosto (mes en el cual se produce un aumento del incumplimiento del SLA, según se ha mostrado en el apartado 2.4), se obtiene un cumplimiento del SLA del 89.5032802249297%. Este aumento del cumplimiento, aproximadamente del 1.70%, se puede explicar en base a diversos factores, siendo uno de ellos una peor asignación de las incidencias a los técnicos por encontrarse de vacaciones los coordinadores más experimentados o los técnicos óptimos. En base a esta información, se plantea como hipótesis que como consecuencia de la implantación del modelo se podrá obtener un aumento del SLA de hasta el 1%, por una asignación óptima de las incidencias teniendo en cuenta la menor afectación de las vacaciones de los coordinadores.

  Este aumento de cumplimiento de SLA tiene como consecuencias directas un menor número de penalizaciones y un aumento del tiempo disponible para poder ejecutar otras incidencias. En relación a ambos, se parte de información de Indea en relación al coste de las penalizaciones (aproximadamente 20 euros) y el beneficio obtenido por cada nueva incidencia gestionada (2 euros).

- **Aumento de asignaciones por soporte del sistema SonaR en el proceso**

  Se estima que los coordinadores realizarán un 40% más de asignaciones gracias al soporte proporcionado por SonaR en el proceso. Según se avance en la implantación del modelo, se podrá prescindir de coordinadores cuando el porcentaje de su tiempo supere el 100% con respecto al volumen de incidencias gestionadas.

4.2.2. Hipótesis en relación a los costes

- **Costes equipo desarrollo**

  Se estima la siguiente dedicación del equipo de trabajo

  - Fases de desarrollo de software: equipo de cuatro personas, trabajando una media de 1,5 horas 3 días a la semana
  - Mantenimiento del modelo: 5% de la dedicación necesaria en las fases de desarrollo de software. El mantenimiento del modelo es continuo desde su puesta en marcha.

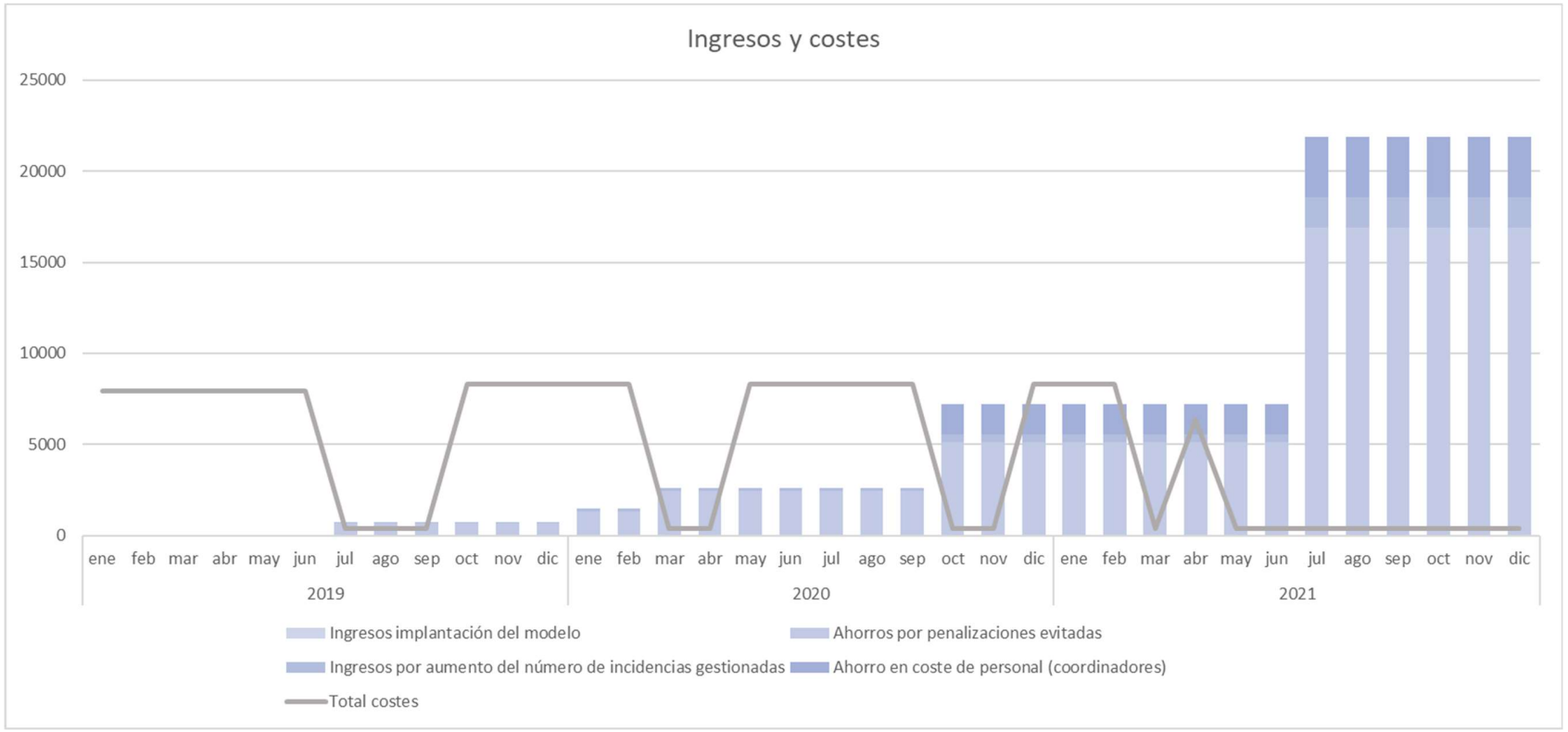
  Se establece un coste de 30 €/hora, en base a estimaciones de mercado.

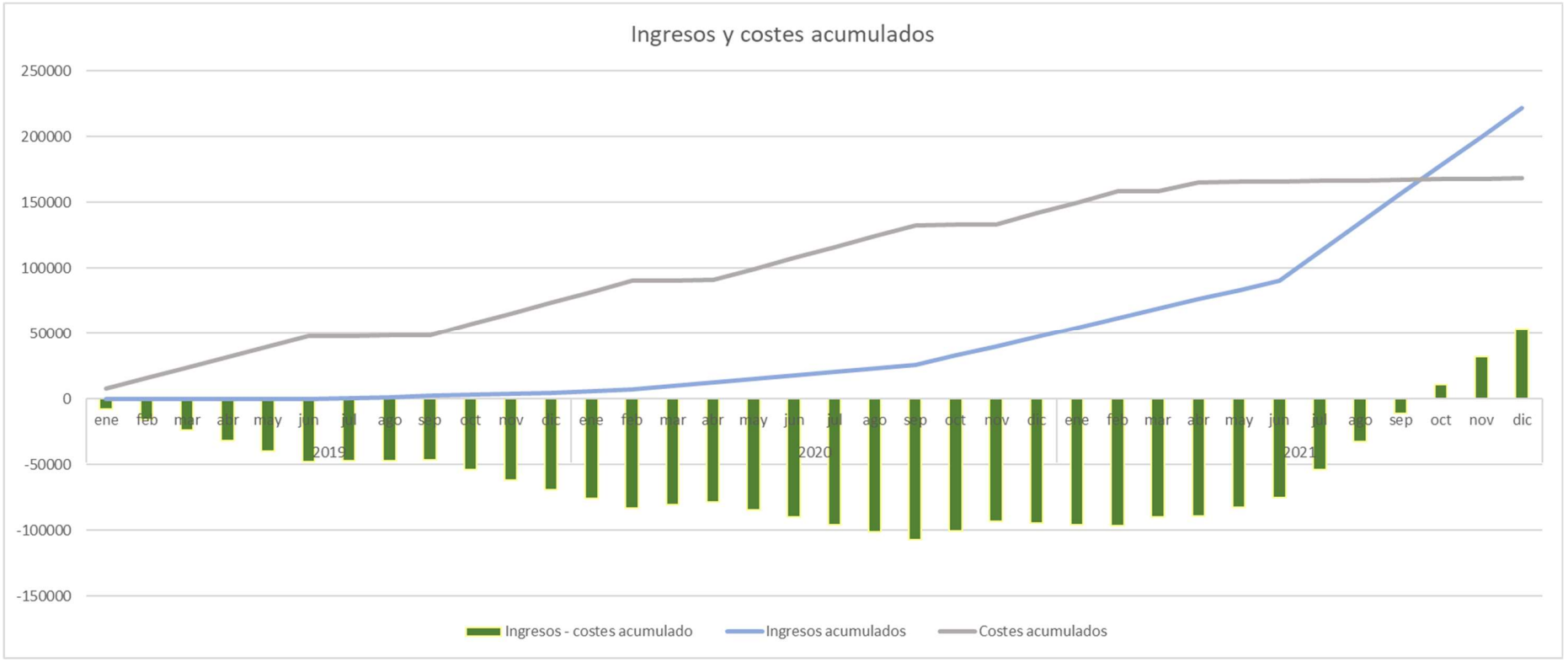
- **Costes infraestructura**

  Únicamente se contempla un refuerzo puntual en infraestructura de almacenamiento y de servidores blade de refuerzo para procesamiento. Se estima un coste de 6.000 € a mediados del tercer año de ejecución. Este coste incluye los servicios de instalación y garantía.

4.3. Business case

En el presente apartado se recogen de forma gráfica los principales resultados obtenidos, así como el resumen de los principales indicadores económicos. Se puede obtener el detalle desglosado en el Excel adjunto a la presente memoria.




Ingresos y costes acumulados
250000
200000
150000
100000
50000
0
-50000
-100000
-150000
ene feb mar abr may jun jul ago sep oct nov dic
2019
ene feb mar abr may jun jul ago sep oct nov dic
2020
ene feb mar abr may jun jul ago sep oct nov dic
2021
Ingresos - costes acumulado
Ingresos acumulados
Costes acumulados

	2019	2020	2021	2019-2021
	Total	Total	Total	Total
Ingresos implantación del modelo				
Ahorros por penalizaciones evitadas	3.999,60 €	35.107,60 €	131.986,80 €	171.094,00 €
Ingresos por aumento del número de incidencias gestionadas	396,00 €	2.816,00 €	12.672,00 €	15.884,00 €
Ahorro en coste de personal (coordinadores)	-00 €	5.000,00 €	30.000,00 €	35.000,00 €
Total ingresos	**4.395,60 €**	42.923,60 €	174.658,80 €	221.978,00 €
Costes implantación del modelo				
Costes equipo desarrollo	73.656,00 €	68.112,00 €	20.592,00 €	162.360,00 €
Costes infraestructura	-00 €	-00 €	6.000,00 €	6.000,00 €
Total costes	**73.656,00 €**	68.112,00 €	26.592,00 €	168.360,00 €
Ingresos acumulados	15.384,60 €	258.140,20 €	1.438.957,80 €	1.712.482,60 €
Costes acumulados	507.276,00 €	1.342.440,00 €	1.968.264,00 €	3.817.980,00 €
Ingresos - costes	**-69.260,40 €**	-25.188,40 €	148.066,80 €	53.618,00 €

Tasa de descuento para el VAR	**15%**
VAR	20.796,42 €
TIR	29%
PAYBACK	34 meses

5. Conclusiones y trabajos futuros

A lo largo de los próximos años, Indea se enfrenta a un reto ambicioso: mejorar su eficiencia en las operaciones para continuar su crecimiento. Son varios los clientes que han manifestado su intención de contar con Indea en nuevas zonas geográficas y ampliar su volumen de negocio. Sin embargo, la falta de automatización de algunos procesos de sus operaciones está impidiendo un mayor control y eficiencia de las mismas. Tal es el caso de la asignación de órdenes de trabajo a los técnicos de campo, tarea que se resuelve a día de hoy gracias a un número reducido de profesionales, que toman decisiones sin una base analítica consistente. De estos coordinadores depende Indea en gran medida, de tal modo que ante sus ausencias o cambios la operación de la compañía se ve afectada, aumentándose el número de incumplimientos por asignaciones incorrectas, y limitando este crecimiento de la empresa. Con el objetivo de automatizar y optimizar este proceso en mente, los equipos de Indea y SonaR acuerdan un proyecto piloto centrado en la operativa en Barcelona, la provincia con mayor número de órdenes de trabajo y técnicos, y limitado al cliente Ingénico, uno de los que generan mayor volumen de trabajo para la empresa.

El piloto desarrollado en el presente proyecto muestra resultados muy esperanzadores para Indea. Gracias a la aplicación de técnicas analíticas y de inteligencia artificial sobre los datos recogidos por la empresa, se comprueba que es posible aportar información de gran valor para soportar y optimizar el proceso de asignación. Para ello, el equipo de SonaR ha seguido un proceso iterativo de depuración de datos, identificación de necesidades, planteamiento de diferentes aproximaciones, así como pruebas iterativas y evaluación de diferentes modelos, para obtener los mejores resultados para el piloto Indea. Los detalles del proceso seguido y decisiones tomadas se han presentado a lo largo del presente documento y se resumen a continuación:

En primer lugar, en el apartado 2 se presenta el proceso de investigación y análisis exploratorio de los datos. En él se perseguía depurar la información disponible, identificar conjuntamente con la empresa necesidades de nueva información, y limpiar el dataset para una correcta aplicación en el modelo.

Adicionalmente, en dicho apartado también se presenta el análisis de las variables disponibles realizado con el objetivo de prescindir de ellas o no en los modelos de inteligencia artificial. De este modo, se ha calculado la aportación de valor predictivo sobre la variable a predecir. De por sí, este proceso ha resultado de gran valor para Indea, ya que al realizar la analítica de las distintas variables, se han detectado patrones e información de interés para la empresa, como por ejemplo los técnicos con mayor incumplimiento de SLA en función del equipamiento a instalar.

Finalmente, como parte del proceso de investigación se ha trabajado con la empresa la creación de nuevas variables que recojan información relevante en el proceso de asignación de incidencias, tales como la distancia a recorrer por el técnico, el número de incidencias abiertas de cada uno de ellos, o un modelado de su experiencia, compuesto a partir de su experiencia de trabajo en clientes, equipos particulares y cumplimiento de SLA. Para una aproximación se ha trabajado adicionalmente en un modelo de experiencia general.

Una vez obtenidos los datasets optimizados, el equipo de SonaR procedió con el entrenamiento de diferentes modelos en base a varias aproximaciones, progresivamente:

- En primera instancia, utilizando entrenamiento de modelos supervisados a partir de la información disponible de Indea, se plantea un modelo de predicción del cumplimiento del SLA de cada técnico ante la llegada de una nueva orden de trabajo. Se entrenaron diversos modelos, obteniendo buenos resultados (*score* superior al 85% con árboles de decisión y Random Forest, por ejemplo). Este modelo debe entenderse como un apoyo a la asignación de las tareas, ya que únicamente ayuda a ver si los técnicos cumplirían o no el SLA, y no ayuda a la asignación automática de técnicos; como consecuencia de ello, se plantearon y desarrollaron modelos adicionales
- En segunda instancia, se planteó el uso de algoritmos no supervisados con el objetivo de identificar clusters de pertenencia de los técnicos, obteniéndose coeficientes de Silhouette superiores a 0.50 con 11 clusters (número calculado mediante el método del codo). Sin embargo, dado que el objetivo del modelo es ayudar a los coordinadores a hacer una asignación óptima de recursos, se decide descartar este modelo como solución al problema descrito en el proyecto, siendo planteado únicamente a modo complementario.
- El tercer modelo planteado se basa en la predicción del rango de la experiencia de los técnicos. Este modelo permite realizar una asignación previa de técnicos, que será utilizada por el coordinador como base de su trabajo. El modelo pretende predecir la experiencia global necesaria para cubrir una nueva orden de trabajo, y clasifica a los técnicos en escalones en base a rangos de experiencia. Para este modelo se obtienen resultados que aportan valor tanto con árboles de decisión como con Random Forest.
- Finalmente, se plantea un modelo de predicción del tiempo de resolución de las incidencias por parte de cada técnico como criterio para decidir cuál es el más apropiado. Partiendo del hecho de que el momento de cierre de la incidencia no es exacto, se plantea la predicción de tramos o horario (franjas de tiempo). Para ello, se siguen procesos iterativos de entrenamiento de

tramos progresivamente más pequeños; lógicamente, el valor predictivo de los modelos se reduce según se exige una mayor exactitud del número de horas (franjas más reducidas), siendo preciso llegar a un compromiso. Se opta por una división en tramos con sentido para el negocio, que tiene en cuenta los tramos relacionados con los dos principales tipos de ordenes de trabajo de Indea: incidencias (SLA de 24 horas) y nuevas instalaciones (SLA de 48 horas).

Para este planteamiento, se han obtenido resultados con score superior al 50% con modelos de árboles de decisión y *Random Forest Classifier*. El análisis del comportamiento de los modelos arroja resultados muy relevantes para Indea: los principales determinantes para predecir la franja de horas en las que se realizará la resolución de una incidencia son comunes a todos los técnicos, tales como el tipo de incidencia, la fecha en la que acontece, o el equipo objeto de la incidencia; adicionalmente, influyen en la decisión otras específicas de cada técnico, tales como el número de incidencias abiertas o la experiencia. Como resultado, al utilizar los modelos con objeto de predicción la mayoría de los técnicos caen en los mismos escalones / tramos horarios de resolución, y solo en el caso del algoritmo Random Forest es capaz de discernir entre los técnicos teniendo en cuenta las características específicas de cada uno de ellos.

Como próximos pasos, a corto plazo se llevará a cabo la puesta en producción del modelo e integración con el aplicativo Pez, según lo acordado con el Director de Sistemas de Información de Indea. A partir de ese momento, será preciso un mantenimiento del sistema mediante reentrenamientos periódicos que permitan adecuar los modelos a las nuevas realidades que vayan surgiendo en la operativa de la compañía.

Una vez puesto en producción el proyecto piloto, se plantea un proceso de expansión del mismo. Por un lado, es preciso seguir mejorando los modelos tecnológicos; en ese sentido, la aplicación de redes neuronales puede suponer un incremento sustancial en el rendimiento de los modelos, según los primeros análisis ya efectuados, y será la primera línea de investigación del equipo de SonaR a partir de este momento. Por otro lado, es preciso adecuar los modelos a más clientes y más zonas geográficas, realizando las adecuaciones que sean precisas a cada contexto.

Anexo I: Entrevistas

a) Notas entrevista 14 /11/2018

Mantenida por equipo de SónaR con Diego Piedrahita vía videoconferencia

Posibilidad de realización de un modelo de Machine Learning para la → **GESTIÓN DE RECURSOS**

Cuando hablamos de recursos, hablamos principalmente de recursos humanos, los **técnicos de calle**

La **gestión de recursos** implica la planificación y movilización de dichos recursos para solucionar problemas reportados por los clientes de la empresa. Actualmente, la gestión de recursos se realiza de forma manual, a través de personas que se denominan **gestores/coordinadores**

Estos gestores/coordinadores, realización la asignación de recursos, dependiendo de una cantidad de **medidores e indicadores** que será necesario tener en cuenta para el desarrollo del modelo de ML.

INDEA trabaja dando **servicios de campo** y haciendo **instalaciones de telecomunicaciones** a través de un pool de **técnicos de calle**

Bolsas de trabajo → De dos tipos:
- Muy específico (Empresas y Casas)
- Instalaciones de Antenas y radio

INDEA cuenta con datos históricos de todos los trabajos realizados, estos trabajos se conocen como **tickets o incidencias**

Diego está interesado en que desarrolle una herramienta para asignar recursos de la mejor manera posible, de forma automática (sin intervención humana) basada en la experiencia de los recursos (que es algo que se mide o se puede construir, a través de varias variables del histórico de datos)

b) Notas entrevista 21/11/2018

Mantenida por equipo de SónaR con Diego Piedrahita vía videoconferencia

Recursos humanos → 200 técnicos de calle

Cuentan con una **plataforma WEB para centralizar las órdenes de trabajo**

2 Modelos de ordenes de trabajo → Día anterior (ejecutarlas de la mejor manera para KPI OK) / Semanal

Ingénico, es uno de los más grandes clientes de INDEA → **las órdenes entran día a día**

Para Ingénico, cuentan con datos históricos desde 2015

Ingénico:

- Las órdenes llegan día a día
- Los técnicos están zonificados
- Llegan cerca de 600 órdenes diarias
- Cuando llega una orden, el coordinador de técnicos realiza la asignación
- Los coordinadores están en Madrid o Barcelona

Vodafone → es otro cliente de INDEA → cuentan con entre 70-80 técnicos para ellos → sólo trabajan en Palma de Mallorca

Volviendo al caso Ingénico…

- Equipos a instalar especificados en el 98% de los casos (caso instalaciones)
- Se realizan también órdenes de trabajo de tipo Mantenimiento
- Es importante controlar las siguientes variables:
    - Stock de equipos del técnico en un momento X
    - Control de histórico de instalación por zona
    - Stock fijo de equipos por técnico

El negocio de Ingénico, consiste en la instalación de TPVs para sus clientes que son los Bancos, entonces, el flujo sería de la siguiente forma:

Cliente del Banco solicita TPV → Banco solicita TPV → Ingénico ordena la instalación de TPV → INDEA realiza la instalación del TPV

El modelo del equipo es una clave para la asignación de recursos → IMPORTANTE - CLAVE

TIP IMPORTANTE: Ingénico vende al banco una cantidad <X> de equipos y los envía a INDEA de acuerdo a la movilidad, luego INDEA distribuye esos equipos entre sus técnicos

El control de Stock es una clave para la asignación de recursos → IMPORTANTE - CLAVE

- Almacén en Barcelona (zona norte)
- Almacén en Madrid
- Almacén en Valencia
- Almacén en Murcia
- Almacén en Palma de Mallorca

PROBLEMA: No hay equipos para cumplir las órdenes a tiempo (48 hrs instalaciones)(24 hrs avaerías)

Para el cumplimiento de KPIs (98% o más):

- Mtto en menos de 24 hrs → **KPIs – IMPORTANTE**
- Instalación en menos de 48 hrs

Asignarle a un técnico una orden, sólo si la puede hacer, esto es:

- Si tiene el equipo
- Dependiendo de la experiencia (que cambia en el tiempo) – No hay control de esto por parte de los coordinadores

La experiencia de los técnicos (cantidad de instalaciones realizadas y modelos que es capaz de instalar) es una clave para la asignación de recursos → IMPORTANTE (*)- CLAVE

(*) Campo a crear en la BD

A veces, muchas órdenes en marcha en el mismo sitio, y los coordinadores no se dan cuenta de ello, es una clave para la asignación de recursos → IMPORTANTE - CLAVE

Existen **bancos diferenciados**, esto sucede por requisito del banco y tienen un acuerdo diferente con Ingénico, un caso de esto es **La Caixa**, que tiene **un KPI de 6hrs para la resolución de incidencias**

Técnicos	98% de Vodafone son propios
	50% de Ingénico son propios

Los técnicos que son autónomos trabajan con bolsas de trabajo, es decir, se les asignan unas incidencias y ellos las resuelven en el tiempo que puedan

INDEA cuenta con un **APP en el móvil** → La incidencia llega al técnico → si la rechaza tiene que ser vía telefónica

Priorizara los técnicos propios debería ser clave para la asignación de recursos → IMPORTANTE - CLAVE

Los técnicos, no pueden contactar con la gente del negocio, es una **NORMA**, y los horarios de los negocios son importantes, si no está la persona cuando el técnico va, implica una parada de reloj y se cobra visita (**pendiente de confirmar**)

El horario del negocio es clave para la asignación de recursos → IMPORTANTE - CLAVE

La intención del modelo es:

- Que no se dependa del coordinador
- Optimizar los recursos
- Reducir el número de técnicos, o no incrementar el número de técnicos ante la creciente entrada de incidencias

KPIs

- 99% Caixa y 98% otros bancos
- Están actualmente entre el 97-98% de cumplimiento, pero les cuesta mucho esfuerzo

Datos para detectar trampas de coordinadores o técnicos:

- A veces utilizan fotos de Google (INDEA tiene guardadas estas fotos en la BD) para hacer paradas de reloj, y hay fotos iguales en varias incidencias, esto **genera vergüenza y pérdida de confianza por parte de Ingénico**
- En caso Vodafone, a veces son los técnicos los que generan las averías.

Identificar técnicos que generan averías o suben fotos falsas → OJO

La gestión logística es un problema – Identificar la ruta más adecuada para repartir modelos – actualmente usan QlickSense – El histórico es clave

Los equipos llegan semanalmente al almacén

Cuando hay una avería (MTTO) → se recoge equipo → se envía a reparación

RMA → recogida y envio de equipos

Estimar subida en cantidad de equipos con averías, si está relacionado a un S/N específico → OJO

Es importante que **no haya paradas por rotura de stock por muchas fallas** – hablar con el director de logística

Rotura de Stock **es clave para la asignación de recursos → IMPORTANTE - CLAVE**

El valor agregado de la solución a desarrollar es la **rapidez en la planificación** (25K instalaciones en 25 días hábiles)

VDF (aunque no entra dentro del scope del proyecto) – Trabajan en Baleares – Pelea por el contrato cada 3 años en base a un ranking que mantiene VDF – hay 14 KPIs

También en caso VDF – cada técnico tiene un extra distinto, por lo tanto la experiencia es difícil de valorar.

PARA EL CASO INGENICO INDEA ES EL UNICO PROVEEDOR

INDEA cuenta con una aplicación (El Pez) que se conecta a la BD Ingénico y trae las órdenes abiertas y pendientes de resolución, también cuenta con una API puente desarrollada para Android que se conecta a El Pez para que los técnicos vean las órdenes.

Cada 10 minutos, el robot, lee las órdenes abiertas y las carga al sistema.

c) Notas entrevista 27/11/2018

Mantenida por equipo de SónaR con Diego Piedrahita vía videoconferencia

Excel de Datos que nos mandó → Datos 2018 – 104.000 entradas – cada visita (sin contar La Caixa) – el cliente es Ingénico.

Orden de Trabajo == Evento

Site == Comercio

Cada vez que el robot se trae los datos, se los trae en formato csv separado con ; → El robot trabaja basado en GETs y POSTs HTTP

El Pez → muestra una matriz por día (la visualización de incidencias con técnicos que nos mostró Diego al compartir pantalla)

INDEA tiene 2 BDs → Firebird (Ingénico)

Oracle

OJO → Es posible meter BD con datos enriquecidos para alimentar el modelo

NOTA IMPORTANTE → Firebird no soporta los datos que tienen y están migrando a Oracle (entiendo yo)

Algunos campos importantes en la info de El Pez:

- Notas Preactuación: Campo txt con la información de la incidencia

- Datos Finalización: Campo txt con la información de cierre

IMPORTANTE → Ver el esquema relacional de las tablas de la BD Firebird

Al cerrar la Incidencia → se meten las notas de finalización en El Pez y se cargan las fotos, mediante una APP que usa el técnico.

IMPORTANTE → Los técnicos tienen un manual para el cierre de las incidencias (Pedirlo a Diego)

Desde la APP se cierran automáticamente en el sistema Pez, hay algunas incidencias que necesitan revisión y otras no.

CESTRACK → Es la tool de Ingénico para el manejo de incidencias

En los datos enviado, el **campo ESTADO_EVENTO es clave para el análisis de tipo de incidencia → IMPORTANTE**

- ANULADO EN PEZ → per no en CESTRACK
- RESUELTO TECNICO → está hecho, pero aún el técnico no ha introducido los datos de cierre en PEZ
- FINALIZADA DELEGACIÓN → el técnico ya lo hizo
- FINALIZADO → Revisado por el coordinador
- SIN RESOLVER → Puede estar incidenciado o no → si está incidenciado implica paradas de tiempo

IMPORTANTE → El modelo debe enfocarse en los casos que están en ESTADO_EVENTO "SIN RESOLVER"

Otros campos importantes (todos los campos fecha en: año-mes-día-hr-min-seg):

- Fecha Entrada → Fecha de creación del evento
- Fecha Planificada → La Fecha en la que dice el coordinador que se va a hacer
- Fecha Tope → algoritmo para el cálculo de fecha límite para no romper SLA
- Fecha Tope – Resuelta → si el valor es positivo está dentro del SLA y si es negativa no
- Horas para cumplir SLA → Campo IMPORTANTE

Tipo de eventos que los técnicos saben hacer (en realidad se traduce como tipo de clientes con los que trabajan), es una clave para la asignación de recursos → IMPORTANTE - CLAVE

Otros campos:

- Latitud y longitud → del comercio que tienen que atender
- NIS → 0 o 1 (dependiendo de si es incidenciado o no, pero tiene más tela)
- Tipo de Evento → I (instalación), M (mantenimiento) o R (recogida)
- El equipo → el modelo se encuentra en el campo Notas de Preactuación, y a veces el técnico no lo conoce (*)
- Horario del Comercio → es un dato a tener en cuenta para la asignación y viene en texto en el campo Notas de Preactuación pero sin formato definido (*)

(*) Campo a crear en la BD

La tabla a continuación muestra algunos datos aproximados:

Madrid	35% de las incidencias	Barcelona	35% de las incidencias
	150 incidencias/día		
	10 técnicos		

Caixa → 250.000 incidencias año

IMPORTANTE

1 evento > = 1 visita (esta es la realidad)

1 evento = 1 visita (es lo deseable y común)

Directo general == mejor coordinador

Pedir a Diego → Datos brutos, datos de Stock

Anexo II: Código Fuente

El código fuente de los notebooks utilizados para la programación del sistema se anexan de forma separada al presente documento.

Anexo III: Informe financiero completo de Indea (fuente: SABI)

Indea Ingenieria De Aplicaciones Sociedad Limitada

46980 PATERNA (VALENCIA, ESPANA)
Empresa privada
El Global Ultimate Owner de esta participada es MR RAFAEL MARTINEZ LUNA

Código NIF B97412175
Fecha últimas cuentas 31/12/2016

Información de contacto

AVENIDA LEONARDO VINCI (PQ. TECNOLOGICO), 18
46980 PATERNA
VALENCIA
ESPANA

Teléfono +34 96/1516201
Fax +34 96/1505733
Dirección web www.indeaingenieria.com

Información legal & tipo cuentas

Forma jurídica	Sociedad limitada	**Ultimo año disponible**	31/12/2016
Forma jurídica detallada	Sociedad limitada	**Años disponibles**	13
Capital social (EUR)	3.006	**Cuentas disponibles**	No Consolidadas
Fecha constitución	09/01/2004		
Estado	Activa		
Estado detallado	Activa		
Director ejecutivo	Don Rafael Martinez Luna		

Información grupo & tamaño (2016)

Ingresos explotación	7.927.950 EUR	**Indicador Independencia BvD**	D
Resultado ejercicio	12.940 EUR	**Empresas en el grupo corporativo**	2
Total activo	4.921.798 EUR	**Núm. accionistas**	1
Número de empleados	100	**Núm. participadas**	1

Clasificación sectorial

Descripción actividad
Servicios de telecomunicaciones.

Código(s) CNAE 2009
Código(s) primario :
6190 - Otras actividades de telecomunicaciones
Código(s) secundario :
6209 - Otros servicios relacionados con las tecnologías de la información y la informática

Código(s) NACE Rev. 2
Código(s) primario :
6190 - Otras actividades de telecomunicaciones
Código(s) secundario :
6209 - Otros servicios relacionados con las tecnologías de la información y la informática

Código(s) CAE Rev.3
Código(s) primario :
6190 - Other telecommunications activities
Código(s) secundario :
6209 - Other information technology and computer service activities

Código(s) US SIC
Código(s) primario :
4899 - Communications services, not elsewhere specified
Código(s) secundario :
7375 - Information retrieval services

7379 - Computer related services, not elsewhere classified

Código(s) IAE
Código(s) primario :
7600 - Telecomunicaciones

NAICS 2017 code(s)
Código(s) primario :
517919 - All Other Telecommunications
Código(s) secundario :
541519 - Other Computer Related Services

Perfil financiero & empleados

Cuentas No Consolidadas	31/12/2016 EUR	31/12/2015 EUR	31/12/2014 EUR	31/12/2013 EUR
	12 meses Aprobado Mixto PGC 2007	12 meses Aprobado Mixto PGC 2007	12 meses Aprobado Mixto PGC 2007	12 meses Abreviado PGC 2007
Ingresos de explotación	7.927.950	7.688.275	6.934.841	6.023.455
Result. ordinarios antes Impuestos	18.442	133.880	504.109	994.970
Resultado del Ejercicio	12.940	93.111	367.557	709.724
Total Activo	4.921.798	3.837.359	3.592.846	3.200.549
Fondos propios	2.432.386	2.656.973	2.563.862	2.200.805
Rentabilidad económica (%)	0,37	3,49	14,03	31,09
Rentabilidad financiera (%)	0,76	5,04	19,66	45,21
Liquidez general	1,34	2,33	2,75	2,96
Endeudamiento (%)	50,58	30,76	28,64	31,24
Número empleados	105	84	86	90

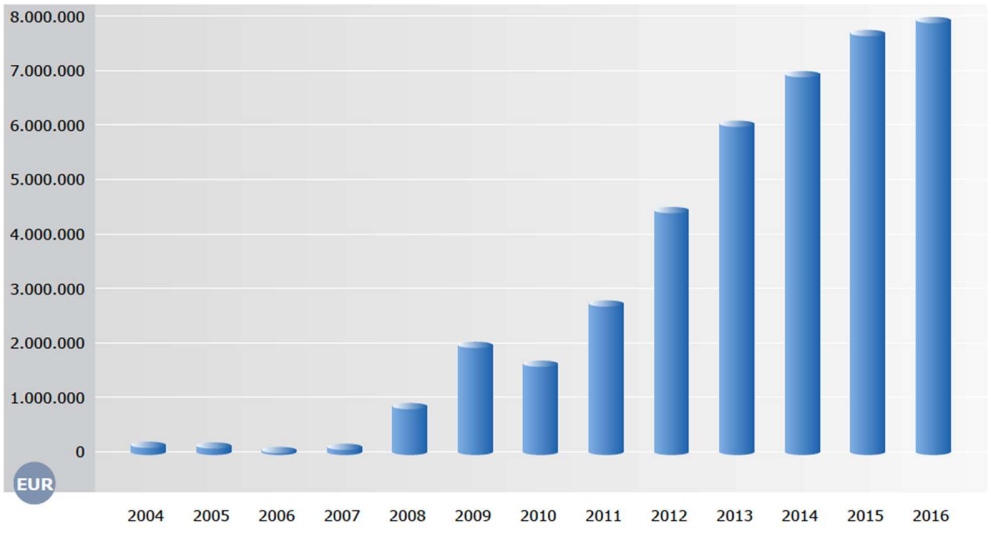
Cuentas No Consolidadas	31/12/2012 EUR	31/12/2011 EUR
	12 meses Pendiente de tratamiento PYME PGC 2007	12 meses Pendiente de tratamiento PYME PGC 2007
Ingresos de explotación	4.438.701	2.724.637
Result. ordinarios antes Impuestos	850.123	362.089
Resultado del Ejercicio	610.086	268.462
Total Activo	2.559.616	1.789.987
Fondos propios	1.491.080	880.995
Rentabilidad económica (%)	33,21	20,23
Rentabilidad financiera (%)	57,01	41,10
Liquidez general	2,13	1,82
Endeudamiento (%)	41,75	50,78
Número empleados	50	22

Cuentas No Consolidadas	31/12/2010 EUR	31/12/2009 EUR	31/12/2008 EUR
	12 meses Pendiente de tratamiento PYME PGC 2007	12 meses Pendiente de tratamiento PYME PGC 2007	12 meses PYME PGC 2007
Ingresos de explotación	1.618.569	1.965.000	840.254
Result. ordinarios antes Impuestos	271.801	406.962	183.570
Resultado del Ejercicio	196.271	290.883	137.873
Total Activo	1.416.728	1.408.912	633.531
Fondos propios	612.532	432.462	140.822
Rentabilidad económica (%)	19,19	28,88	28,98
Rentabilidad financiera (%)	44,37	94,10	130,36
Liquidez general	1,85	1,46	1,18
Endeudamiento (%)	56,76	69,31	77,77
Número empleados	4	4	4

Cuentas No Consolidadas	31/12/2007 EUR	31/12/2006 EUR	31/12/2005 EUR
	12 meses Pendiente de tratamiento Abreviado	12 meses Abreviado	12 meses Pendiente de tratamiento Abreviado
Ingresos de explotación	94.342	34.022	117.176
Result. ordinarios antes Impuestos	36.332	-9.603	-16.739
Resultado del Ejercicio	24.134	-9.603	-24.074
Total Activo	179.613	58.225	40.117
Fondos propios	-6.486	-30.620	-21.018
Rentabilidad económica (%)	20,23	-16,49	-41,72
Rentabilidad financiera (%)	-560,17	31,36	79,64
Liquidez general	1,01	0,48	0,31
Endeudamiento (%)	103,61	152,59	152,39
Número empleados	n.d.	1	1

Cuentas No Consolidadas	31/12/2004 EUR
	12 meses Pendiente de tratamiento Abreviado
Ingresos de explotación	130.057
Result. ordinarios antes Impuestos	118
Resultado del Ejercicio	83
Total Activo	64.558
Fondos propios	3.089
Rentabilidad económica (%)	0,18
Rentabilidad financiera (%)	3,82
Liquidez general	0,57
Endeudamiento (%)	95,22
Número empleados	1

Evolución de una variable clave: Ingresos de explotación (2004 - 2016)



Formato Global

Cuentas No Consolidadas	31/12/2016 EUR	31/12/2015 EUR	31/12/2014 EUR	31/12/2013 EUR
	12 meses Aprobado Mixto PGC 2007	12 meses Aprobado Mixto PGC 2007	12 meses Aprobado Mixto PGC 2007	12 meses Abreviado PGC 2007
Balance de situación				
Inmovilizado	1.618.795	1.144.043	928.881	524.423
Inmovilizado inmaterial	2.602	3.425	5.664	n.d.
Inmovilizado material	184.658	233.171	260.451	227.654
Otros activos fijos	1.431.535	907.447	662.767	296.769
Activo circulante	3.303.002	2.693.316	2.663.965	2.676.126
Existencias	518.944	129.701	44.095	n.d.
Deudores	2.408.547	1.297.584	1.415.197	1.758.523
Otros activos líquidos	375.512	1.266.031	1.204.672	917.603
Tesorería	211.645	1.158.162	1.096.804	769.734
Total activo	4.921.798	3.837.359	3.592.846	3.200.549
Fondos propios	2.432.386	2.656.973	2.563.862	2.200.805
Capital suscrito	3.006	3.006	3.006	3.006
Otros fondos propios	2.429.380	2.653.967	2.560.856	2.197.799
Pasivo fijo	17.110	26.130	61.117	96.007
Acreedores a L. P.	17.110	26.130	61.117	96.007
Otros pasivos fijos	0	0	0	0
Provisiones	n.d.	n.d.	n.d.	n.d.
Pasivo líquido	2.472.302	1.154.256	967.866	903.737
Deudas financieras	1.261.320	30.376	88.466	128.411
Acreedores comerciales	1.103	16.483	2.018	7.989
Otros pasivos líquidos	1.209.880	1.107.397	877.382	767.338
Total pasivo y capital propio	4.921.798	3.837.359	3.592.846	3.200.549
Fondo de maniobra	2.926.388	1.410.803	1.457.274	1.750.534
Número empleados	105	84	86	90
Cuentas de pérdidas y ganancias				
Ingresos de explotación	7.927.950	7.688.275	6.934.841	6.023.455
Importe neto Cifra de Ventas	7.917.710	7.684.437	6.928.443	6.011.395
Consumo de mercaderías y de materias	n.d.	n.d.	n.d.	n.d.
Resultado bruto	n.d.	n.d.	n.d.	n.d.
Otros gastos de explotación	n.d.	n.d.	n.d.	n.d.
Resultado Explotación	45.450	234.029	563.908	1.033.690
Ingresos financieros	0	0	0	0
Gastos financieros	27.008	100.149	59.799	38.719
Resultado financiero	-27.008	-100.149	-59.799	-38.719
Result. ordinarios antes Impuestos	18.442	133.880	504.109	994.970
Impuestos sobre sociedades	5.502	40.769	136.552	285.246
Resultado Actividades Ordinarias	12.940	93.111	367.557	709.724
Ingresos extraordinarios	n.d.	n.d.	n.d.	n.d.
Gastos extraordinarios	n.d.	n.d.	n.d.	n.d.
Resultados actividades extraordinarias	n.d.	n.d.	n.d.	n.d.
Resultado del Ejercicio	12.940	93.111	367.557	709.724
Materiales	3.788.882	3.160.902	2.335.455	1.389.175
Gastos de personal	2.948.971	2.954.560	2.543.427	2.216.317
Dotaciones para amortiz. de inmovil.	60.534	61.849	65.176	52.552
Other operating items	-1.084.114	-1.276.935	-1.426.876	-1.331.721
Gastos financieros y gastos asimilados	27.008	37.797	48.047	38.719
Cash flow	73.474	154.960	432.734	762.277
Valor agregado	3.054.955	3.188.086	3.160.759	3.302.559
EBIT	45.450	234.029	563.908	1.033.690

Cuentas No Consolidadas	31/12/2012 EUR 12 meses Pendiente de tratamiento PYME PGC 2007	31/12/2011 EUR 12 meses Pendiente de tratamiento PYME PGC 2007
Balance de situación		
Inmovilizado	742.573	651.207
Inmovilizado inmaterial	n.d.	n.d.
Inmovilizado material	278.331	227.051
Otros activos fijos	464.242	424.156
Activo circulante	1.817.043	1.138.780
Existencias	n.d.	n.d.
Deudores	1.217.690	738.843
Otros activos líquidos	599.354	399.937
Tesorería	468.485	286.844
Total activo	2.559.616	1.789.987
Fondos propios	1.491.080	880.995
Capital suscrito	3.006	3.006
Otros fondos propios	1.488.074	877.989
Pasivo fijo	214.283	283.290
Acreedores a L. P.	214.283	283.290
Otros pasivos fijos	0	0
Provisiones	n.d.	n.d.
Pasivo líquido	854.252	625.703
Deudas financieras	131.445	221.847
Acreedores comerciales	468	21.239
Otros pasivos líquidos	722.339	382.618
Total pasivo y capital propio	2.559.616	1.789.987
Fondo de maniobra	1.217.221	717.605
Número empleados	50	22
Cuentas de pérdidas y ganancias		
Ingresos de explotación	4.438.701	2.724.637
Importe neto Cifra de Ventas	4.435.115	2.723.453
Consumo de mercaderías y de materias	n.d.	n.d.
Resultado bruto	n.d.	n.d.
Otros gastos de explotación	n.d.	n.d.
Resultado Explotación	886.822	408.591
Ingresos financieros	0	0
Gastos financieros	36.700	46.502
Resultado financiero	-36.700	-46.502
Result. ordinarios antes Impuestos	850.123	362.089
Impuestos sobre sociedades	240.037	93.627
Resultado Actividades Ordinarias	610.086	268.462
Ingresos extraordinarios	n.d.	n.d.
Gastos extraordinarios	n.d.	n.d.
Resultados actividades extraordinarias	n.d.	n.d.
Resultado del Ejercicio	610.086	268.462
Materiales	804.041	633.842
Gastos de personal	1.624.113	996.336
Dotaciones para amortiz. de inmovil.	58.062	41.297
Other operating items	-1.065.662	-644.572
Gastos financieros y gastos asimilados	36.700	46.502
Cash flow	668.148	309.759
Valor agregado	2.568.998	1.446.223
EBIT	886.822	408.591

Cuentas No Consolidadas	31/12/2010 EUR	31/12/2009 EUR	31/12/2008 EUR
	12 meses Pendiente de tratamiento PYME PGC 2007	12 meses Pendiente de tratamiento PYME PGC 2007	12 meses PYME PGC 2007
Balance de situación			
Inmovilizado	604.817	451.625	51.817
Inmovilizado inmaterial	n.d.	192	342
Inmovilizado material	196.170	43.044	51.474
Otros activos fijos	408.647	408.389	n.d.
Activo circulante	811.912	957.287	581.715
Existencias	n.d.	n.d.	n.d.
Deudores	629.435	662.634	363.320
Otros activos líquidos	182.476	294.654	218.395
Tesorería	79.383	223.404	218.395
Total activo	1.416.728	1.408.912	633.531
Fondos propios	612.532	432.462	140.822
Capital suscrito	3.006	3.006	3.006
Otros fondos propios	609.526	429.456	137.816
Pasivo fijo	364.538	318.765	67
Acreedores a L. P.	364.538	318.765	67
Otros pasivos fijos	0	0	0
Provisiones	n.d.	n.d.	n.d.
Pasivo líquido	439.658	657.685	492.642
Deudas financieras	154.492	203.040	n.d.
Acreedores comerciales	56.521	n.d.	131.865
Otros pasivos líquidos	228.645	454.645	360.777
Total pasivo y capital propio	1.416.728	1.408.912	633.531
Fondo de maniobra	572.914	662.634	231.455
Número empleados	4	4	4
Cuentas de pérdidas y ganancias			
Ingresos de explotación	1.618.569	1.965.000	840.254
Importe neto Cifra de Ventas	1.610.712	1.964.766	840.254
Consumo de mercaderías y de materias	n.d.	n.d.	n.d.
Resultado bruto	n.d.	n.d.	n.d.
Otros gastos de explotación	n.d.	n.d.	n.d.
Resultado Explotación	312.349	419.126	187.519
Ingresos financieros	0	1	0
Gastos financieros	40.549	12.164	3.949
Resultado financiero	-40.549	-12.164	-3.949
Result. ordinarios antes Impuestos	271.801	406.962	183.570
Impuestos sobre sociedades	75.530	116.078	45.697
Resultado Actividades Ordinarias	196.271	290.883	137.873
Ingresos extraordinarios	n.d.	n.d.	n.d.
Gastos extraordinarios	n.d.	n.d.	n.d.
Resultados actividades extraordinarias	n.d.	n.d.	n.d.
Resultado del Ejercicio	196.271	290.883	137.873
Materiales	293.371	633.101	352.798
Gastos de personal	568.410	515.269	124.483
Dotaciones para amortiz. de inmovil.	21.598	9.581	8.545
Other operating items	-422.841	-387.923	-166.909
Gastos financieros y gastos asimilados	40.549	12.164	1.974
Cash flow	217.868	300.464	146.417
Valor agregado	902.357	943.976	318.572
EBIT	312.349	419.126	187.519

Cuentas No Consolidadas	31/12/2007 EUR	31/12/2006 EUR	31/12/2005 EUR
	12 meses Pendiente de tratamiento Abreviado	12 meses Abreviado	12 meses Pendiente de tratamiento Abreviado
Balance de situación			
Inmovilizado	n.d.	24.023	27.994
Inmovilizado inmaterial	n.d.	107	128
Inmovilizado material	n.d.	23.916	27.865
Otros activos fijos	n.d.	0	0
Activo circulante	179.613	34.202	12.124
Existencias	n.d.	n.d.	5.140
Deudores	109.436	1.738	5.540
Otros activos líquidos	70.177	32.464	1.444
Tesorería	70.177	32.464	1.444
Total activo	179.613	58.225	40.117
Fondos propios	-6.486	-30.620	-21.018
Capital suscrito	3.006	3.006	3.006
Otros fondos propios	-9.492	-33.626	-24.024
Pasivo fijo	8.028	18.190	22.269
Acreedores a L. P.	8.028	18.190	22.269
Otros pasivos fijos	n.d.	n.d.	n.d.
Provisiones	n.d.	n.d.	n.d.
Pasivo líquido	178.071	70.656	38.866
Deudas financieras	n.d.	n.d.	n.d.
Acreedores comerciales	n.d.	n.d.	n.d.
Otros pasivos líquidos	178.071	70.656	38.866
Total pasivo y capital propio	179.613	58.225	40.117
Fondo de maniobra	109.436	1.738	10.680
Número empleados	n.d.	1	1
Cuentas de pérdidas y ganancias			
Ingresos de explotación	94.342	34.022	117.176
Importe neto Cifra de Ventas	94.342	34.022	117.176
Consumo de mercaderías y de materias	n.d.	n.d.	n.d.
Resultado bruto	n.d.	n.d.	n.d.
Otros gastos de explotación	n.d.	n.d.	n.d.
Resultado Explotación	37.119	-8.825	-14.884
Ingresos financieros	n.d.	n.d.	n.d.
Gastos financieros	788	778	1.855
Resultado financiero	-788	-778	-1.855
Result. ordinarios antes Impuestos	36.332	-9.603	-16.739
Impuestos sobre sociedades	5.017	n.d.	-68
Resultado Actividades Ordinarias	31.315	-9.603	-16.671
Ingresos extraordinarios	n.d.	n.d.	594
Gastos extraordinarios	7.180	n.d.	7.997
Resultados actividades extraordinarias	-7.180	n.d.	-7.403
Resultado del Ejercicio	24.134	-9.603	-24.074
Materiales	51.034	18.647	76.150
Gastos de personal	4.058	13.171	26.341
Dotaciones para amortiz. de inmovil.	n.d.	3.971	9.054
Other operating items	n.d.	n.d.	n.d.
Gastos financieros y gastos asimilados	788	778	1.855
Cash flow	24.134	-5.632	-15.020
Valor agregado	33.997	8.316	13.109
EBIT	37.119	-8.825	-14.884

Cuentas No Consolidadas	31/12/2004 EUR 12 meses Pendiente de tratamiento Abreviado
Balance de situación	
Inmovilizado	47.277
Inmovilizado inmaterial	6.128
Inmovilizado material	41.149
Otros activos fijos	0
Activo circulante	17.281
Existencias	10.980
Deudores	1.326
Otros activos líquidos	4.975
Tesorería	4.556
Total activo	64.558
Fondos propios	3.089
Capital suscrito	3.006
Otros fondos propios	83
Pasivo fijo	31.377
Acreedores a L. P.	31.377
Otros pasivos fijos	n.d.
Provisiones	n.d.
Pasivo líquido	30.093
Deudas financieras	n.d.
Acreedores comerciales	n.d.
Otros pasivos líquidos	30.093
Total pasivo y capital propio	64.558
Fondo de maniobra	12.307
Número empleados	1
Cuentas de pérdidas y ganancias	
Ingresos de explotación	130.057
Importe neto Cifra de Ventas	130.057
Consumo de mercaderías y de materias	n.d.
Resultado bruto	n.d.
Otros gastos de explotación	n.d.
Resultado Explotación	2.703
Ingresos financieros	n.d.
Gastos financieros	2.585
Resultado financiero	-2.585
Result. ordinarios antes Impuestos	118
Impuestos sobre sociedades	35
Resultado Actividades Ordinarias	83
Ingresos extraordinarios	n.d.
Gastos extraordinarios	n.d.
Resultados actividades extraordinarias	n.d.
Resultado del Ejercicio	83
Materiales	70.456
Gastos de personal	25.537
Dotaciones para amortiz. de inmovil.	11.492
Other operating items	n.d.
Gastos financieros y gastos asimilados	2.585
Cash flow	11.575
Valor agregado	39.731
EBIT	2.703

Cuentas No Consolidadas	31/12/2016 EUR	31/12/2015 EUR	31/12/2014 EUR	31/12/2013 EUR
EBITDA	105.984	295.878	629.084	1.086.242

Cuentas No Consolidadas	31/12/2012 EUR	31/12/2011 EUR
EBITDA	944.885	449.888

Cuentas No Consolidadas	31/12/2010 EUR	31/12/2009 EUR	31/12/2008 EUR
EBITDA	333.947	428.707	196.064

Cuentas No Consolidadas	31/12/2007 EUR	31/12/2006 EUR	31/12/2005 EUR
EBITDA	37.119	-4.854	-5.830

Cuentas No Consolidadas	31/12/2004 EUR
EBITDA	14.195

Ratios formato global

Cuentas No Consolidadas	31/12/2016 EUR	31/12/2015 EUR	31/12/2014 EUR	31/12/2013 EUR
	12 meses Aprobado Mixto PGC 2007	12 meses Aprobado Mixto PGC 2007	12 meses Aprobado Mixto PGC 2007	12 meses Abreviado PGC 2007
A. Rentabilidad				
Rentabilidad sobre recursos propios (%)	0,76	5,04	19,66	45,21
Rentabilidad sobre capital empleado (%)	1,86	6,40	21,03	45,01
Rentabilidad sobre el activo total (%)	0,37	3,49	14,03	31,09
Margen de beneficio (%)	0,23	1,74	7,27	16,52
B. Operaciones				
Rotación de activos netos	3,24	2,87	2,64	2,62
Ratio de cobertura de intereses	1,68	6,19	11,74	26,70
Rotación de las existencias	15,28	59,28	157,27	n.s.
Período de cobro (días)	109	61	73	105
Período de crédito (días)	0	1	0	0
C. Estructura				
Ratio de solvencia	1,34	2,33	2,75	2,96
Ratio de liquidez	1,13	2,22	2,71	2,96
Ratios de autonomía financiera a medio y largo plazo	142,17	101,68	41,95	22,92
Coeficiente de solvencia (%)	49,42	69,24	71,36	68,76
Apalancamiento (%)	52,56	2,13	5,83	10,20
D. Por empleado				
Beneficio por empleado	0	2	6	11
Ingresos de explotación por empleado	76	92	81	67
Costes de los trabajadores / Ingresos de explotación (%)	37,20	38,43	36,68	36,79
Coste medio de los empleados	28	35	30	25
Recursos propios por empleado	23	32	30	24
Capital circulante por empleado	28	17	17	19
Total activos por empleado	47	46	42	36

Cuentas No Consolidadas	31/12/2012 EUR 12 meses Pendiente de tratamiento PYME PGC 2007	31/12/2011 EUR 12 meses Pendiente de tratamiento PYME PGC 2007
A. Rentabilidad		
Rentabilidad sobre recursos propios (%)	57,01	41,10
Rentabilidad sobre capital empleado (%)	52,00	35,09
Rentabilidad sobre el activo total (%)	33,21	20,23
Margen de beneficio (%)	19,15	13,29
B. Operaciones		
Rotación de activos netos	2,60	2,34
Ratio de cobertura de intereses	24,16	8,79
Rotación de las existencias	n.s.	n.s.
Período de cobro (días)	99	98
Período de crédito (días)	0	3
C. Estructura		
Ratio de solvencia	2,13	1,82
Ratio de liquidez	2,13	1,82
Ratios de autonomía financiera a medio y largo plazo	6,96	3,11
Coeficiente de solvencia (%)	58,25	49,22
Apalancamiento (%)	23,19	57,34
D. Por empleado		
Beneficio por empleado	17	16
Ingresos de explotación por empleado	89	124
Costes de los trabajadores / Ingresos de explotación (%)	36,59	36,57
Coste medio de los empleados	32	45
Recursos propios por empleado	30	40
Capital circulante por empleado	24	33
Total activos por empleado	51	81

Cuentas No Consolidadas	31/12/2010 EUR	31/12/2009 EUR	31/12/2008 EUR
	12 meses Pendiente de tratamiento PYME PGC 2007	12 meses Pendiente de tratamiento PYME PGC 2007	12 meses PYME PGC 2007
A. Rentabilidad			
Rentabilidad sobre recursos propios (%)	44,37	94,10	130,36
Rentabilidad sobre capital empleado (%)	31,97	55,79	131,69
Rentabilidad sobre el activo total (%)	19,19	28,88	28,98
Margen de beneficio (%)	16,79	20,71	21,85
B. Operaciones			
Rotación de activos netos	1,66	2,62	5,96
Ratio de cobertura de intereses	7,70	34,46	94,99
Rotación de las existencias	n.s.	n.s.	n.s.
Período de cobro (días)	140	121	156
Período de crédito (días)	13	0	56
C. Estructura			
Ratio de solvencia	1,85	1,46	1,18
Ratio de liquidez	1,85	1,46	1,18
Ratios de autonomía financiera a medio y largo plazo	1,68	1,36	2.101,83
Coeficiente de solvencia (%)	43,24	30,69	22,23
Apalancamiento (%)	84,73	120,66	0,05
D. Por empleado			
Beneficio por empleado	68	102	46
Ingresos de explotación por empleado	405	491	210
Costes de los trabajadores / Ingresos de explotación (%)	35,12	26,22	14,81
Coste medio de los empleados	142	129	31
Recursos propios por empleado	153	108	35
Capital circulante por empleado	143	166	58
Total activos por empleado	354	352	158

Cuentas No Consolidadas	31/12/2007 EUR 12 meses Pendiente de tratamiento Abreviado	31/12/2006 EUR 12 meses Abreviado	31/12/2005 EUR 12 meses Pendiente de tratamiento Abreviado
A. Rentabilidad			
Rentabilidad sobre recursos propios (%)	-560,17	31,36	79,64
Rentabilidad sobre capital empleado (%)	2.406,68	71,00	-1.188,95
Rentabilidad sobre el activo total (%)	20,23	-16,49	-41,72
Margen de beneficio (%)	38,51	-28,22	-14,29
B. Operaciones			
Rotación de activos netos	61,17	-2,74	93,60
Ratio de cobertura de intereses	47,14	-11,35	-8,02
Rotación de las existencias	n.s.	n.s.	22,80
Período de cobro (días)	418	18	17
Período de crédito (días)	0	0	0
C. Estructura			
Ratio de solvencia	1,01	0,48	0,31
Ratio de liquidez	1,01	0,48	0,18
Ratios de autonomía financiera a medio y largo plazo	-0,81	-1,68	-0,94
Coeficiente de solvencia (%)	-3,61	-52,59	-52,39
Apalancamiento (%)	-123,78	-59,41	-105,96
D. Por empleado			
Beneficio por empleado	n.s.	-10	-17
Ingresos de explotación por empleado	n.s.	34	117
Costes de los trabajadores / Ingresos de explotación (%)	4,30	38,71	22,48
Coste medio de los empleados	n.s.	13	26
Recursos propios por empleado	n.s.	-31	-21
Capital circulante por empleado	n.s.	2	11
Total activos por empleado	n.s.	58	40

Cuentas No Consolidadas	31/12/2004 EUR
	12 meses Pendiente de tratamiento Abreviado
A. Rentabilidad	
Rentabilidad sobre recursos propios (%)	3,82
Rentabilidad sobre capital empleado (%)	7,84
Rentabilidad sobre el activo total (%)	0,18
Margen de beneficio (%)	0,09
B. Operaciones	
Rotación de activos netos	3,77
Ratio de cobertura de intereses	1,05
Rotación de las existencias	11,84
Período de cobro (días)	4
Período de crédito (días)	0
C. Estructura	
Ratio de solvencia	0,57
Ratio de liquidez	0,21
Ratios de autonomía financiera a medio y largo plazo	0,10
Coeficiente de solvencia (%)	4,78
Apalancamiento (%)	1.015,88
D. Por empleado	
Beneficio por empleado	0
Ingresos de explotación por empleado	130
Costes de los trabajadores / Ingresos de explotación (%)	19,63
Coste medio de los empleados	26
Recursos propios por empleado	3
Capital circulante por empleado	12
Total activos por empleado	65

Informaciones bursátiles

No hay datos bursátiles para esta empresa.

Administradores / contactos actuales

Juntas y comités

= *también accionista*

	Nombre	Título original de la función	Comité	Fuente
1.	Don Rafael Martinez Luna P002537434	- Administrador Único (desde 09/10/2007)	BoD	IN
			Informa (recibido 02/03/2019)	

Administración y personal

= *también accionista*

	Nombre	Título original de la función	Departamento	Fuente
1.	Don Rafael Martinez Luna P002537434	- Director General (desde 29/04/2015)	SenMan	IN
			Informa (recibido 02/03/2019)	
		- Director Financiero (desde 29/04/2015)	FinAcc	IN
			Informa (recibido 16/02/2019)	
		- Director Comercial (desde 29/04/2015)	Sales	IN
			Informa (recibido 16/02/2019)	
2.	Doña Lorena Moreno Munguia P347456923	- Director Técnico (desde 28/04/2016)	R&D	IN
			Informa (recibido 16/02/2019)	

3.	Doña Maria Jose Gascon P347456922	-Director de Calidad (desde 28/04/2016)	Qual, OthDep	IN Informa (recibido 16/02/2019)
4.	Don Diego Priedrahita P347456921	-Director de Informática (desde 28/04/2016)	IT&IS	IN Informa (recibido 16/02/2019)

Auditores de Cuentas y Bancos

Auditor: Betea Espana Auditores S.L.P. (Última fecha nombramiento: 16/02/2018)
Provalencia Consultores S.L. (Última fecha nombramiento: 31/12/2015)

Estado auditoría por año:

2016 cuenta:
Fecha de cierre: 31/12/2016
Opinión auditores: Aprobado
Nombre del auditor: PROVALENCIA CONSULTORES S.L.
Código de auditor: S1601

2015 cuenta:
Fecha de cierre: 31/12/2015
Opinión auditores: Aprobado
Nombre del auditor: PROVALENCIA CONSULTORES S.L.
Código de auditor: S1601

2014 cuenta:
Fecha de cierre: 31/12/2014
Opinión auditores: Aprobado
Nombre del auditor: PROVALENCIA CONSULTORES SRL

2013 cuenta:
Fecha de cierre: 31/12/2013

2012 cuenta:
Fecha de cierre: 31/12/2012
Opinión auditores: Pendiente de tratamiento

2011 cuenta:
Fecha de cierre: 31/12/2011
Opinión auditores: Pendiente de tratamiento

2010 cuenta:
Fecha de cierre: 31/12/2010
Opinión auditores: Pendiente de tratamiento

2009 cuenta:
Fecha de cierre: 31/12/2009
Opinión auditores: Pendiente de tratamiento

2008 cuenta:
Fecha de cierre: 31/12/2008

2007 cuenta:
Fecha de cierre: 31/12/2007
Opinión auditores: Pendiente de tratamiento

2006 cuenta:
Fecha de cierre: 31/12/2006

2005 cuenta:
Fecha de cierre: 31/12/2005
Opinión auditores: Pendiente de tratamiento

2004 cuenta:
Fecha de cierre: 31/12/2004
Opinión auditores: Pendiente de tratamiento

Bancos:

BANKINTER
NOVO B SUCURSAL EN ESPAÑA

Accionistas actuales

Filtro actual: Sin filtrar

	Nombre del accionista	País	Tipo	Accionistas Directo (%)	Accionistas Total (%)	Fuente	Fecha de la info.	Vari-ación	Información empresa Ingreso Operacional (mll EUR)*	Información empresa Empleados
1.	MR RAFAEL MARTINEZ LUNA	n.d.	I	100,00	100,00	IN	02/2019		-	-

=También un director
* = *Para una compañía de seguros el valor correspondiente es el Gross Premium Written y para un banco es el Operating Income (memo)*

Participadas actuales

Filtro actual: Sin filtrar

Las empresas subrayadas y presentadas en azul y en negrita están disponibles en **SABI**

	Nombre participada	País	Accionistas Directo (%)	Accionistas Total (%)	Nivel de acc.	Estado	Fuente	Fecha de la info.	Vari-ación	Información empresa Ingreso Operacional (mll EUR)*	Información empresa Empleados
1.	ICONO INDEA UTE	ES	40,00	n.d.	1	-	IN	02/2019		n.d.	n.d.

* = *Para una compañía de seguros el valor correspondiente es el Gross Premium Written y para un banco es el Operating Income (memo)*