

Con la colaboración de la Escuela de Organización Industrial.



Desarrollado como proyecto final del Curso Ejecutivo en Big Data & Business Analytics 2014 – 2015.

Grupo de trabajo A, formado por:



Antonio Alonso Piñero



Ricardo Arteaga Vioque



Iván Cabañas Polo



Alberto Fernández Sánchez

Índice

1	Introducción.....	8
2	Problema de negocio	9
2.1	Situación actual.....	10
2.2	Comparativa de España frente a Europa.....	10
2.2.1	Estudios e iniciativas sobre la reincidencia.....	12
2.2.2	Fuentes estadísticas	15
2.3	El reto: Objetivo y oportunidad.....	16
2.4	Motivación y viabilidad	19
2.4.1	Innovación.....	19
2.4.2	Viabilidad técnica	20
2.4.3	Viabilidad económica	20
3	¿Quiénes son nuestros clientes?	21
3.1	Descripción del marco cliente.....	21
3.2	Descripción de la SGIIPP.....	23
3.3	Contactos y entrevistas	29
3.4	Otros posibles clientes	30
4	¿Quiénes somos?	32
4.1	Definición instrumental de la empresa.....	32
4.2	Misión, visión y valores	32
4.2.1	Misión	33
4.2.2	Visión.....	33
4.2.3	Valores.....	33
4.3	Plan estratégico.....	33
4.3.1	Mapa estratégico	33
4.3.2	Indicadores clave (KPIs)	35
4.3.3	Debilidades, amenazas, fortalezas y oportunidades	38
4.4	Modelo de negocio.....	40
5	Solución	42
5.1	Definición del producto	42
5.2	Análisis de las fuentes de datos	43
5.2.1	Expediente Judicial	44

5.2.2	Expediente Penitenciario	45
5.2.3	Contenido publicado en Internet y redes sociales	49
5.3	Modelo lógico	49
5.4	Arquitectura	50
5.5	Funcionalidad de la aplicación y perfiles de usuarios.....	54
5.5.1	Perfiles de usuarios.....	54
5.5.2	Funcionalidad de la aplicación.....	55
5.6	Modelo predictivo.....	56
5.6.1	Inventario de Factores de Riesgo e Intervención en Prisiones (IFRIP)	57
5.6.2	Análisis descriptivo de las características de la muestra....	58
5.6.3	Relación entre el análisis de la reincidencia delictiva y otros grupos de variables.....	58
5.6.4	Análisis del Inventario de Factores de Riesgo e Intervención en Prisiones (IFRIP).....	59
5.6.5	Modelos explicativos de la reincidencia delictiva	63
5.7	Motor de recomendaciones.....	66
5.8	Visualización.....	67
6	Plan de proyecto.....	71
6.1	Metodología	71
6.2	Implantación	74
6.2.1	Planificación de tareas.....	75
6.2.2	Fases de implantación de Predictive Data Justice	75
6.2.3	Riesgos.....	77
7	Prueba de concepto	78
7.1	Entorno de las pruebas	79
7.2	Nube de Tags.....	80
7.3	Dendograma y clustering sobre texto	81
7.4	Relaciones de reincidencia	84
8	Business case.....	85
8.1	Inversión y costes	86
8.2	Rentabilidad económica	87
9	Conclusiones.....	89

10	Mejoras	89
11	Bibliografía y referencias	90
12	Glosario	92
13	Anexos	94
13.1	Centros Penitenciarios	94
13.2	Criterios de evaluación de riesgos	95
13.3	Mock-ups de la aplicación visual	97
13.4	Código fuente en R	101

Índice de figuras

Figura 2-1. Pirámide de conocimiento.....	10
Figura 2-2. Mapa de calor de la tasa de población carcelaria	11
Figura 2-3. Impacto sobre los procesos judiciales y de reinserción en un individuo	17
Figura 3-1. Organigrama de la SGIIPP.....	24
Figura 3-2. Relaciones entre la SGIIPP y sus entidades colaboradoras.	25
Figura 3-3. Tasa de población reclusa en Sudamérica	31
Figura 4-1. DAFO	38
Figura 4-2. Modelo de negocio.....	41
Figura 5-1. Logotipo del producto	43
Figura 5-2. Modelo lógico	50
Figura 5-3. Arquitectura física.....	51
Figura 5-4. Tiempo de regresión logística entre Hadoop y Spark.....	52
Figura 5-5. Stack tecnológico de la arquitectura física.....	53
Figura 5-6. Curva ROC	62
Figura 5-7. Modelo explicativo de la reincidencia delictiva	65
Figura 5-8. Login de usuario	67
Figura 5-9. Buscador de recluso.....	68
Figura 5-10. Pestaña del histórico judicial	69
Figura 5-11. Análisis de riesgo de reincidencia.....	69
Figura 6-1. Inicio del proyecto.....	72
Figura 6-2. Análisis y planificación.....	73
Figura 6-3. Diseño y construcción	73
Figura 6-4. Validación y despliegue.....	74
Figura 6-5. Soporte y mantenimiento	74

Figura 6-6. Cronograma	75
Figura 6-7. Distribución territorial Centros Penitenciarios.....	76
Figura 7-1. Ejemplo de RStudio	79
Figura 7-2. Resultado de la ejecución de la nube de tags	81
Figura 7-3. Modificación del atributo Sparsity en la Matriz	82
Figura 7-4. Dendrograma con 10 clústeres ejecutado sobre sentencias	83
Figura 7-5. Dendrograma con 10 clústeres ejecutado sobre sentencias	84
Figura 8-1. Flujo neto efectivo.....	88
Figura 8-2. Curva ROC de reincidencia	88
Figura 12-1. Centros Penitenciarios	94

Índice de tablas

Tabla 2-1. Estadísticas penitenciarias de España frente a Europa.....	12
Tabla 2-2. Estadísticas sobre reincidencia general y específica en España.....	12
Tabla 2-3. Estadísticas sobre reincidencia específica por droga en España.....	13
Tabla 2-4. Estadísticas sobre reincidencia en España clasificadas	13
Tabla 2-5. Oportunidades del proyecto	18
Tabla 2-6. Retos del proyecto.....	18
Tabla 2-7. Objetivos generales del proyecto.....	18
Tabla 3-1. Funciones de la SGIIPP	23
Tabla 3-2. Entidades colaboradoras con la SGIIPP.....	25
Tabla 4-1. Valores de Predictive Data.....	33
Tabla 4-3. Indicadores clave de desempeño (KPIs).....	35
Tabla 5-1. Estructura y tipos de datos contenidos en el fichero SISPE.	46
Tabla 5-2. Reincidencia general frente a tiempo de condena.....	59
Tabla 5-3. Coordenadas de la curva ROC para la reincidencia delictiva	60
Tabla 5-4. Correlaciones entre cada uno de los factores del IFRIP y la reincidencia delictiva	63
Tabla 5-1. Plan de implantación	77
Tabla 5-2. Riesgos iniciales identificados en el proyecto.....	78
Tabla 8-1. Reducción tasa de reincidencia	86
Tabla 8-2. Costes de implantación Predictive Data Justice.....	86
Tabla 8-3. Costes de la Administración en programas de reinserción...	87
Tabla 8-4. Indicadores de rentabilidad	87
Tabla 8-5. Resumen flujo neto.....	87
Tabla 10-1. Revistas electrónicas de interés	90

Tabla 10-2. Observatorios y centros de documentación	90
Tabla 10-3. Fuentes estadísticas.....	91

1 Introducción

La presente memoria tiene como principal objetivo complementar al resumen ejecutivo del proyecto **Predictive Data Justice**, aportando detalles que son necesarios para una completa y profunda comprensión del propósito de negocio propuesto.

Incluye una versión detallada de **Predictive Data Justice**, necesario para su completa comprensión y alcance del producto, desde el punto de vista técnico y económico.

2 Problema de negocio

Existe la creencia generalizada de que los delincuentes que se reinsertan en la sociedad presentan una alta probabilidad de que vuelvan a cometer un delito. Sin embargo, la distribución de la reincidencia delictiva, o en otras palabras, de la comisión de nuevos delitos, es en realidad muy heterogénea y varía en función del delito cometido.

Consecuentemente, es evidente la importancia que tiene analizar los factores que están implicados en el riesgo de reincidencia delictiva, y no sólo para la elaboración de normas jurídico-penales mejores, sino también para la propia ejecución penal y el cumplimiento de la pena privativa de libertad.

Bajo este contexto nace nuestro proyecto **Predictive Data Justice**, una solución para la **evaluación del riesgo de reincidencia delictiva** entre los delincuentes, analizando de una forma conveniente la influencia de todos los posibles factores de riesgo que confluyan en cada caso, para establecer una serie de recomendaciones basadas en posibles tratamientos, que logren una completa rehabilitación del delincuente.

De este modo, se trata de emplear de una forma innovadora, las capacidades de conocimiento proporcionadas por **Big Data**, para construir una medición objetiva, científica y con carácter universal del riesgo de reincidencia delictiva, basándose en el análisis de los datos que en combinación con el instinto y la experiencia, permita potenciar una adopción mucho más eficiente en la toma de decisiones.

No cabe ninguna duda, que nuestra solución sirve para mejorar la aplicabilidad de las políticas de prevención delictiva, optimizar el proceso penitenciario, con una mejor asignación de las actividades terapéutico-asistenciales, directamente dirigidas y orientadas hacia la consecución efectiva de la reeducación y **reinserción social del delincuente**, y aportar un gran **valor social** para la ciudadanía, incrementando el estado de bienestar y la seguridad ciudadana.



Figura 2-1. Pirámide de conocimiento

2.1 Situación actual

Presentamos en este capítulo algunos datos de interés, que han sido recopilados para nuestro proyecto durante los últimos meses, con la clara finalidad de dar a conocer el estado en el que se encuentran estadísticamente el delito, la justicia y las penas en España.

Todas las referencias extraídas se encuentran detalladas en el [Capítulo 11 Bibliografía y referencias](#).

2.2 Comparativa de España frente a Europa

Desde un punto de vista estadístico, España lidera el ranking de los países de Europa Occidental en cuanto a número de reclusos y centros penitenciarios se refiere.

Mediante el siguiente mapa, se ilustra la tasa de población carcelaria por cada 100.000 habitantes. Estos datos se han extraído del estudio “Realidad y Política Penitenciarias”, realizado por el Observatorio de la Delincuencia (ODA) del Instituto Andaluz Interuniversitario de Málaga, en el año 2012.

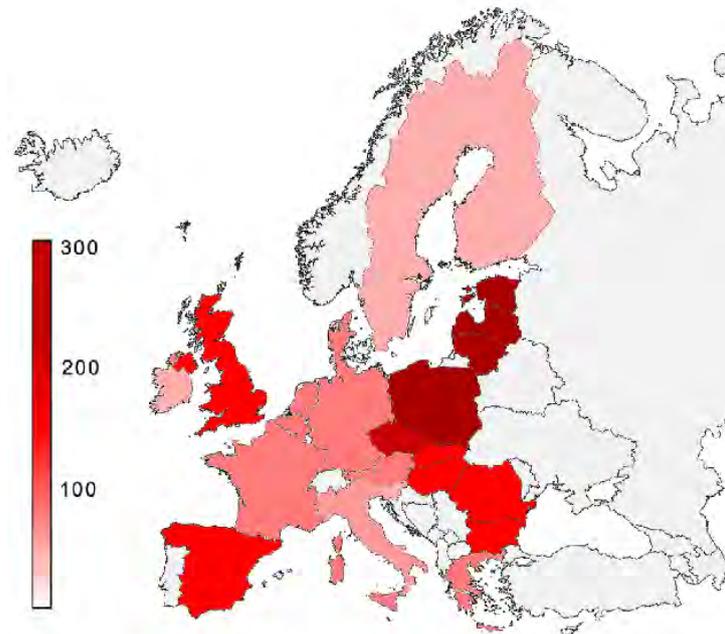


Figura 2-2. Mapa de calor de la tasa de población carcelaria

España posee una tasa de reclusos¹ situada en **153 por cada 100.000 habitantes**, más del doble de la Unión Europea, situada en 70. Contrasta fuertemente con otras tasas como los 84,6 de Alemania, los 111,6 de Italia y los 117 por cada 100.000 habitantes en Francia.

Además paradójicamente España es uno de los países con cifras de criminalidad más baja de Europa: 46,1 delitos por cada mil habitantes, frente a los 62,8 de media europea.

¹ La **tasa de reclusos, población penitenciaria o densidad de población penitenciaria** siempre se mide en unidades de **número de reclusos por cada 100.000 habitantes**

Tabla 2-1. Estadísticas penitenciarias de España frente a Europa

	España	Europa
Coste mensual por recluso	1950€	3090€
Tiempo medio de estancia en centros penitenciarios	19 meses	10 meses
Porcentaje de reclusos por robo	30,70%	12,90%
Tasa de reclusos por 100.000 habitantes	153/100.000	70/100.000
Posición en el ranking de número de reclusos	4º Posición	
Posición en el ranking de condenas más extensas	6ª Posición	

2.2.1 Estudios e iniciativas sobre la reincidencia

El Consejo Europeo urge a adoptar medidas, entre las que destacan la incentivación de políticas de reeducación y de reinserción social, como por ejemplo pueden ser los trabajos comunitarios, como forma principal y alternativa a la privación de la libertad.

Según un estudio de la **Fundación Atenea**, dos de cada tres reclusos vuelve a cometer un delito cuando queda puesto en libertad, lo que revela una elevada tasa reincidencia entre los reclusos españoles.

Tabla 2-2. Estadísticas sobre reincidencia general y específica en España

	España
Reincidencia general.	50%
Reincidencia específica sobre el mismo delito.	65%

Además, al comparar los niveles de reincidencia por tipo de delito, se ve un fuerte aumento en el caso de los hurtos "ya que pasa de 71% en la primera medición a un 90% en la segunda. El robo sube de un 41% a un 46%, y otros delitos contra la

propiedad (distintos al hurto y a los robos con violencia), aumentan del 26% al 44%".

Tabla 2-3. Estadísticas sobre reincidencia específica por droga en España

	España
Porcentaje de segunda reincidencia por droga.	50%
Porcentaje de reincidentes por droga.	90%
Porcentaje de reincidencia en los dos primeros años.	80%

Los estudios advierten que existe una especialización en el delito: quienes ingresan del centro penitenciario y retornan por nuevas condenas, vuelven en su mayoría por el mismo delito que habían egresado.

Queda claro que el paso por el centro penitenciario no es efectivo para corregir este tipo de conductas, ya que "el 80% de las reincidencias ocurre los dos primeros años de libertad".

La **Fundación Atenea** destaca la falta de programas efectivos para tratar a los reclusos adictos a las drogas, ya que "sólo el 20% recibe tratamiento para la desintoxicación".

El colectivo de reclusos con ciertas **adicciones** a sustancias es el más numeroso. Se calcula que el 80% de los reclusos consumían drogas antes de ingresar en un centro penitenciario. Este dato puede explicar el hecho de que España, pese a ser uno de los países con menor delincuencia, se encuentra entre los primeros países de Europa por delitos contra la salud pública, por delante de Francia, Italia, Alemania, Bulgaria o Reino Unido.

Tabla 2-4. Estadísticas sobre reincidencia en España clasificadas

	Total	Reincidente	%
Homicidio y formas	395	76	19,24%
Del Aborto	5	1	20,00%
De las lesiones	4.802	534	11,12%

De las lesiones al feto	8	1	12,50%
Relativos a la manipulación genética	3	1	33,33%
Contra la libertad	704	124	17,61%
De las torturas y otros delitos contra la integridad moral	275	39	14,18%
Contra la libertad e indemnidad sexuales	949	119	12,54%
De la omisión del deber de socorro	16	2	12,50%
Contra la intimidad, derecho a la propia imagen e inviolable	76	12	15,79%
Contra el honor	60	4	6,67%
Contra las relaciones familiares	2.242	270	12,04%
Contra el patrimonio y el orden socioeconómico	35.257	12.261	34,78%
Contra la Hacienda pública y la Seguridad Social	206	14	6,80%
Contra los derechos de los trabajadores	396	13	3,28%
Relativos a la ordenación territorial, protección del patrimonio y medio ambiente	131	13	9,92%
Contra la seguridad colectiva	35.266	4.438	12,58%
De las falsedades	2.380	256	10,76%
Contra la Administración Pública	170	14	8,24%
Contra la Administración de Justicia	2.735	1.296	47,39%
Contra la Constitución	892	29	3,25%
Contra el orden público	3.484	673	19,32%

Delitos de traición y contra la paz o independencia	4	0	0,00%
Contra la comunidad internacional	0	0	0,00%
Concurso de delitos	7.083	2.092	29,54%
No consta	308	65	21,10%
Total	97.847	22.347	22,84%

El Consejo Europeo, alarma sobre la masificación de los centros penitenciarios en países como España y Reino Unido, y urge a utilizar vías de castigo alternativas. Entre las recomendaciones para disminuir la población penitenciaria, destacan el empleo de medias alternativas a la prisión como pueden ser la libertad vigilada, los trabajos para la comunidad, la localización permanente y las multas.

El trabajo comunitario, como principal alternativa a la privación de libertad, no debe reducirse a la producción o al simple hecho de estar ocupado: “El principal objetivo debe seguir siendo que quien haya cometido un delito, trabaje en torno al delito y en proyectos sociales relacionados”.

El Consejo Europeo apuesta por utilizar los centros penitenciarios solo como último recurso y por aplicar esas medidas alternativas en las penas más cortas. Entre los países que han apostado por estas vías alternativas destaca Holanda, que en la última década ha conseguido reducir un 22% su población penitenciaria.

2.2.2 Fuentes estadísticas

El análisis de los registros criminales es fundamental en la prevención del delito. Entre otras cosas, porque permite el diseño de políticas y planes de prevención efectivos. En la **Administración General del Estado (AGE)**, este tipo de análisis se ha realizado históricamente mediante herramientas estadísticas descriptivas básicas, considerando fundamentalmente variables y relaciones primarias. Sin embargo, se ha demostrado que la estadística descriptiva clásica no refleja la verdadera interrelación de las variables y por lo tanto, no aporta una solución efectiva al problema real.

Por otra parte, el **Ministerio del Interior (MI)** cuenta con la explotación del Sistema Estadístico de Criminalidad (SEC), donde únicamente se reflejan datos cuantitativos recogidos por las **Fuerzas y Cuerpos de Seguridad del Estado (FyCSE)**.

La **Secretaría General de Instituciones Penitenciarias (SGIIPP)**, elabora sus propios informes estadísticos penitenciarios, que publica en su página web de forma semanal e incluso mensual. En definitiva, todos los análisis estadísticos que se realizan y que hemos expuesto aquí son descriptivos.

2.3 El reto: Objetivo y oportunidad

Muchos de los procesos de la **AGE** son legados y difícilmente mejorables en cuanto a las nuevas tecnologías se refiere. Actualmente, se sigue generando una gran cantidad de información altamente relevante que se continúa transcribiendo en papel. Sin embargo, en estos últimos años, tanto el **MI** como el **MJ** han liderado un importante avance tecnológico en busca de un claro objetivo, como es el de la digitalización de todos sus expedientes, con una consiguiente reducción del gasto en papel.

En concreto, los sistemas de gestión procesal han continuado evolucionando progresivamente, en busca de una mayor interoperabilidad entre los distintos sistemas. Claro ejemplo de ello es el **Expediente Judicial Electrónico (EJE)** que promueve el MJ.

Por tanto, todo ese gran volumen de información generada de una forma continuada por la propia AGE, principalmente por la SGIIPP, el MI y el MJ, no se está aprovechando para extraer conocimiento ni valor aplicable a la toma de decisiones.

Gracias a la irrupción de este nuevo paradigma digital del Big Data, surgen nuevas oportunidades de negocio, que unidas a la aparición de nuevas tecnologías, generan el '*maching*' o simbiosis perfecta entre ambas.

La oportunidad de aprovechar un entorno completamente **digitalizado**, unido al tratamiento masivo de la información existente que proviene de diversas fuentes a priori no interconectadas entre sí, y a la utilización de ciertos modelos analíticos-predictivos, abre una nueva puerta de mercado al mundo de la inteligencia de negocio en la toma de decisiones aplicable dentro del sistema penitenciario, que permita una evaluación objetiva, científica y universal del riesgo de reincidencia delictiva, analizar de una forma conveniente la influencia de todos los posibles factores de riesgo que confluyan en cada caso, para establecer una serie de recomendaciones basadas en posibles tratamientos, que logren una completa reinserción del delincuente.

Este contexto plantea un reto de negocio altamente innovador, que necesita de unos indicadores robustos, que a su vez puedan ofrecer métricas claras y transparentes de la mejora del proceso penitenciario.

Los estudios presentados hasta el momento, revelan que un alto porcentaje de la población reclusa es reincidente. Si a esto unimos además que algunos de los factores más influyentes sobre la reincidencia son perfectamente identificables, hace pensar que utilizando un correcto modelo analítico-predictivo, alimentado con toda la información conocida del recluso, permitirá conocer de una forma mucho más eficaz, cuáles son las acciones que se deben aplicar de forma específica, para lograr un completa rehabilitación del recluso.

El gasto incurrido por la **AGE** en materias de penitenciaría, seguridad ciudadana y justicia, se encuentra conceptualizado dentro de los **Presupuestos Generales del Estado (PGE)**, como un gasto primario, básico y de carácter obligatorio. En este sentido, nuestra principal intención consiste en **proporcionar una mejor adecuación de la partida presupuestaria destinada a la SGIIPP**, mediante la optimización del proceso penitenciario, a través de una mejor asignación de las actividades terapéutico-asistenciales, directamente dirigidas y orientadas hacia la consecución efectiva de la reeducación y **reinserción social** del delincuente, y aportar un gran **valor social** para la ciudadanía.

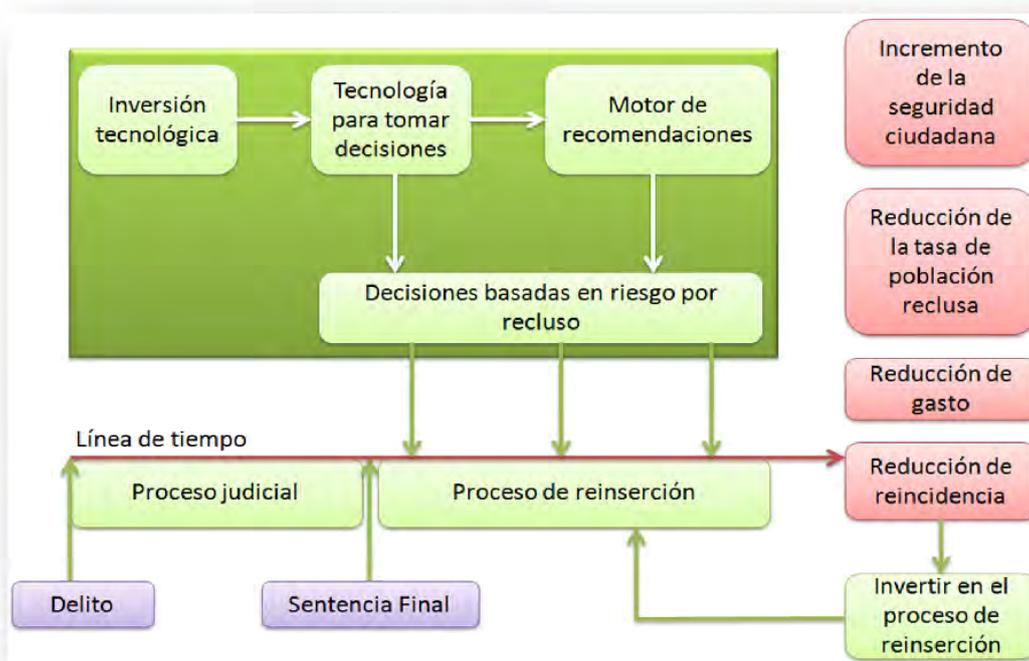


Figura 2-3. Impacto sobre los procesos judiciales y de reinserción en un individuo

Tabla 2-5. Oportunidades del proyecto

Digitalización de la documentación que conlleva un procedimiento judicial y su ejecución penal en el sistema penitenciario. Volumen de datos extenso

Extracción de una información del expediente judicial a través de las sentencias y autos, así como el historial o expediente penitenciarios en los textos.

Las herramientas actuales y estudios estadísticos descriptivos no permiten extraer conclusiones veraces. No se analiza el problema real

Pocas soluciones alternativas a la nuestra, ante la falta de competidores

Tabla 2-6. Retos del proyecto

Implantar un potente análisis de datos utilizando modelos analíticos-predictivos para mejorar la eficiencia en la toma de decisiones, dentro del proceso penitenciario

Focalizar el caso de uso de reincidencia delictiva que mayor impacto posee en la tasa de población reclusa

Transmitir el conocimiento de las nuevas tecnologías basadas en Big Data, poniendo en evidencia tanto sus ventajas como su menor coste computacional

Apertura de un nuevo modelo de negocio altamente innovador que aporte un gran valor a la AGE

Tabla 2-7. Objetivos generales del proyecto

Reducción de la tasa de reclusos reincidentes, descongestionando los centros penitenciarios, tomando como referencia la tasa media de reclusos reincidentes de la Unión Europea

Incremento del gasto medio diario por recluso

Reducción del tiempo medio de estancia en los centros penitenciarios, para equiparlo con la media de la Unión Europea

Transformación del proceso penitenciario para una completa reinserción de los reclusos y su reincorporación a la sociedad

Por lo tanto, hay que admirar la **gran dimensión social** que se pretende abarcar con este proyecto, ya que no solo se trata de brindar una nueva oportunidad a los reclusos de incorporarse a la sociedad, sino también una ventana de tiempo para un cambio de mentalidad, una forma totalmente distinta de tomar decisiones basadas en la inteligencia de los datos, en definitiva una visión mucho más horizontal de la realidad del ser humano.

2.4 Motivación y viabilidad

En este capítulo se analizan las **claves**, que desde el punto de vista de la **innovación, viabilidad técnica y económica**, nos llevaron a decantarnos finalmente por la elaboración y desarrollo del proyecto **Predictive Data Justice**.

2.4.1 Innovación

El **valor principal** que aporta la **inteligencia de negocio**, permite **minimizar los riesgos en la toma de decisiones**. La principal función de la **SGIIPP** es la reinserción en la sociedad de los reclusos que hayan cometido algún tipo de delito. Actualmente, esto se realiza en base al instinto y a la experiencia previa, pero sin tener en cuenta datos analíticos ni predictivos, ya que **los procesos de reinserción no son medidos ni personalizados**, lo que conlleva a que en la inmensa mayoría de los casos se fracase en el intento de reinserción. Por lo tanto, es una **gran oportunidad** de utilizar el avance de las nuevas tecnologías aplicadas de forma inteligente y personalizada, para mejorar la toma de decisiones en el ámbito penitenciario.



¿Por qué Predictive Data Justice es un proyecto innovador?

- ✓ Aporta una dimensión social.
- ✓ Utiliza la inteligencia de negocio para optimizar y mejorar la toma de decisiones dentro del marco de la legalidad.
- ✓ Emplea nuevas tecnologías que nunca antes se habían aplicado al entorno penitenciario.
- ✓ Aprovecha toda la información histórica relativa tanto al proceso judicial como al proceso penitenciario, gracias a la enorme capacidad que aporta el Big Data en cuanto al almacenamiento masivo de la información y su procesamiento se refiere.

- ✓ Utiliza minería de texto en un entorno jurídico-penitenciario completamente veraz, con una completa ausencia de la ironía.

2.4.2 Viabilidad técnica

Tal y como se ha mencionado al comienzo de este capítulo, el **ingente crecimiento de la información** producida en el proceso judicial y penitenciario, es muy difícil de tratar con el estado del arte de las tecnologías actualmente implantadas en dicho entorno jurídico-penitenciario, aunque siempre es de admirar los continuos esfuerzos que desde la **AGE** se realizan para invertir en las nuevas tecnologías.

La SGIIPP ha invertido parte de su presupuesto anual en el desarrollo de un servicio telemático y sanitario para los reclusos, y de un portal de acceso para el ciudadano, pero sin embargo **carece de un sistema completo de inteligencia de negocio**.

La propia implantación y despliegue de un complejo sistema de inteligencia de negocio, alimentado por el gran volumen de información que actualmente se encuentra disponible, y que proviene de diversas fuentes, parece utópico y poco probable si no se adopta el nuevo paradigma tecnológico de **Big Data**.



¿Qué aporta el Big Data a la solución Predictive Data Justice?

- ✓ Extraer un conocimiento y valor de la información que se encuentra disponible, y que nunca antes se había empleado con ese fin.
- ✓ Proporcionar el desarrollo y la construcción de una solución muy económica, y con una gran capacidad de almacenamiento y de procesamiento.
- ✓ Transversalidad en una solución altamente escalable.
- ✓ Analizar textos previamente digitalizados sin necesidad de estructurarlos ni de almacenarlos en una base de datos relacional.

2.4.3 Viabilidad económica

El hecho de poner en valor la inmensa mayoría de la información que se genera y almacena en el ámbito jurídico-penitenciario, hace que se adopte un nuevo paradigma acerca de la repercusión que el tratamiento de dicha información tendrá sobre mejora en la toma de decisiones para paliar la reincidencia delictiva.

Los estudios estadísticos basados en los factores que influyen sobre la reincidencia delictiva, son claves para comprender su **gran impacto económico**.

La reducción de la reincidencia delictiva determinará cuáles de los gastos llevados a cabo dentro del entorno jurídico-penitenciario, poseen un mayor o menor grado de efectividad entre la población reclusa, con lo que se podrá llegar a **redirigir** de una forma mucho más inteligente, todas aquellas **partidas presupuestarias hacia las medidas que tengan una tasa alta de efectividad en la disolución del comportamiento delictivo de los reclusos.**



¿Por qué Predictive Data Justice es un proyecto económicamente viable?

- ✓ La inversión inicial es relativamente baja con respecto al retorno que se puede llegar a obtener.
- ✓ Al ser una solución transversal altamente escalable, posee un bajo coste de mantenimiento.
- ✓ Proporciona una solución tecnológica de mejora sobre los estudios estadísticos actualmente existentes.
- ✓ Dentro de la AGE, tanto el MJ como el MI, la SGIIPP, y las FyCSE están apostando por la inversión en este tipo paradigma tecnológico del Big Data.

3 ¿Quiénes son nuestros clientes?

3.1 Descripción del marco cliente

Este proyecto queda enmarcado, de una forma general, en el ámbito de la Administración Pública de España (APE), que es un elemento no político del poder ejecutivo, formado por organismos e instituciones, servido por funcionarios que gestionan los intereses, servicios públicos y recursos del Estado.

Los principios generales constitucionales de la organización administrativa, se encuentran desarrollados legalmente por la Ley 6/97, del 14 de Abril, de Organización y Funcionamiento de la Administración General del Estado (LOFAGE, artículos 3 y 4), estableciendo que “**La Administración General del Estado (AGE)** es aquella que extiende sus competencias a todo el territorio español, bajo la dirección del Gobierno. Está formada por una serie de órganos jerárquicamente ordenados y actúa con personalidad jurídica única, desarrollando funciones ejecutivas de carácter administrativo”.

En consecuencia, es de considerar de una manera mucho más específica, que este proyecto pretende dar un servicio de alto valor añadido a la AGE. Por eso y de un modo mucho más concreto y conciso, como entidad dependiente de la AGE, el cliente fundamental y principal al que va dirigido este proyecto es la

Secretaría General de Instituciones Penitenciarias (SGIIPP), dependiente de la Secretaría de Estado de Seguridad del Ministerio del Interior (MI), cuyos fines quedan perfectamente definidos mediante el artículo 25.2 de la Constitución Española de 1978, y el artículo 1 de la Ley Orgánica 1/79, del 26 de Septiembre, General Penitenciaria, fundamentándose en garantizar el cumplimiento de las penas impuestas por los jueces, en la reeducación y reinserción social de los sentenciados a penas y medidas penales privativas de libertad, en la retención y custodia de detenidos, presos y penados, así como en la asistencia y ayuda de los internos y liberados, y en garantizar su integridad.

Así mismo, consideramos que también son clientes e intervinientes de nuestro proyecto todas aquellas entidades colaboradoras con la SGIIPP, como por ejemplo pueden ser las **Organizaciones No Gubernamentales** (ONG's, como la Unión Española de Asociaciones y Entidades de Atención al Drogodependiente – UNAD, Cruz Roja Española, Proyecto Hombre, Caritas Española, etc.), **asociaciones, fundaciones** (como la Fundación Padre Garralda-Horizontes Abiertos, Fundación Diagrama, Fundación Adsis, Fundación ATENEA, etc.), entidades privadas con fines de interés público, y otras entidades de voluntariado social, cuyo papel resulta esencial para la colaboración y participación de la sociedad en la intervención penitenciaria a la hora de conseguir la reinserción y reeducación de las personas condenadas.

Además, las **Fuerzas y Cuerpos de Seguridad del Estado (FyCSE)** se consideran como potenciales clientes de este proyecto, ya que están integrados tanto por el Cuerpo Nacional de Policía (CNP) como por la Guardia Civil (GC), dirigidos respectivamente por la Dirección General de la Policía y por la Dirección General de la Guardia Civil, bajo la dependencia de la Secretaría de Estado de Seguridad del **Ministerio del Interior (MI)**, y cuya regulación esencial reside en el Título II de la Ley Orgánica 2/1986, de Fuerzas y Cuerpos de Seguridad, puesto que se han de encargar del mantenimiento del orden, de proteger el libre ejercicio de los derechos y libertades, de garantizar la seguridad pública y ciudadana, así como de la prevención e investigación del delito.

Por otra parte, es necesario incorporar también como cliente de nuestro proyecto al **Ministerio de Justicia (MJ)**, departamento ministerial que como legislador posee competencias en justicia, y se encarga entre otros de preparar, dirigir, y ejecutar la política del Gobierno de la Nación para el desarrollo del ordenamiento jurídico, especialmente en materia de derecho penal, civil, mercantil y procesal, de la política de organización y apoyo de la Administración de Justicia, de la cooperación jurídica internacional, así como con las comunidades autónomas en

coordinación con los demás departamentos competentes en la materia, y de la asistencia jurídica del Estado.

3.2 Descripción de la SGIIPP

Tal y como se ha comentado en el capítulo anterior, hemos focalizado principalmente el desarrollo de este proyecto sobre la SGIIPP, y también por afinidad, sobre todas aquellas entidades que colaboran con la SGIIPP.

Las funciones principales de la SGIIPP se enumeran en la siguiente tabla:

Tabla 3-1. Funciones de la SGIIPP

Función inspectora sobre los servicios, organismos y centros de la administración penitenciaria

Tramitación de las informaciones reservadas y de los expedientes disciplinarios incoados a los funcionarios y personal laboral destinados en todas sus unidades

La administración y gestión del personal que presta servicio en los centros y servicios dependientes

La propuesta en la planificación y seguimiento del desarrollo del Plan de creación de infraestructuras penitenciarias

La presidencia de la Entidad Estatal de Derecho Público de Trabajo Penitenciario y Formación para el Empleo

La planificación y control de las actividades tendentes al mantenimiento y mejora de la higiene y de la salud en el medio penitenciario, así como la prevención, tratamiento y rehabilitación de drogodependencias

Desde un punto de vista organizativo, la SGIIPP se estructura internamente en siete subestructuras de dirección y de operación relacionadas con la gestión de las penitenciarías.



Figura 3-1. Organigrama de la SGIIPP

El objetivo principal de nuestro proyecto consiste en proveer de un sistema basado en inteligencia de negocio a la SGIIPP y entidades colaboradoras, con el objetivo de reducir el porcentaje de reclusos que cometen reincidencia delictiva, adoptando acciones adecuadas y personalizadas para lograr la rehabilitación y la reinserción de los reclusos, con el mínimo coste repercutido a la sociedad.

A continuación, la siguiente figura representa las relaciones más importantes que se establecen entre nuestro principal cliente (la SGIIPP), y sus entidades colaboradoras.



Figura 3-2. Relaciones entre la SGIIPP y sus entidades colaboradoras.

En la actualidad, la SGIIPP tiene establecidos vínculos de colaboración con todas las asociaciones recogidas en la siguiente tabla:

Tabla 3-2. Entidades colaboradoras con la SGIIPP

<p>UNAD</p>	<p>La Unión de Asociaciones y Entidades de Atención al Drogodependiente, es una organización no gubernamental, sin ánimo de lucro, que interviene en el ámbito de las drogodependencias y los problemas derivados de las mismas</p>
<p>Fundación Padre Garralda-Horizontes Abiertos</p>	<p>Organización no gubernamental que se dedica a la integración social de las personas más necesitadas. Toda su labor está concebida para dar esperanza a las personas más marginadas de la sociedad, a través del desarrollo de diferentes proyectos dirigidos a facilitar su acceso a una vida normalizada</p>

**Cruz Roja
Española**

Institución humanitaria, de carácter voluntario y de interés público, que desarrolla su actividad bajo la protección del Gobierno de España y el Alto Patronazgo de los Reyes de España

**Proyecto
Hombre**

Organización sin ánimo de lucro que trabaja para ayudar a las personas en dificultad o en desventaja social, contribuyendo a que puedan desarrollarse personalmente, recobren su autonomía y se integren en la sociedad, a través de programas y acciones terapéutico-educativas estructuradas según las necesidades de los diferentes perfiles

**Caritas
Española**

Organización oficial de la Iglesia católica en España para la acción caritativa y social, instituida por la Conferencia Episcopal Española

**Fundación
Diagrama**

Entidad que trabaja en la atención de las necesidades de personas vulnerables o en dificultad social, siempre desde la defensa y promoción de los Derechos Humanos

**Fundación
Adsis**

Organización no gubernamental, promovida por el Movimiento de Comunidades Adsis, que trabaja para la promoción integral de personas y grupos empobrecidos, principalmente jóvenes

FEAPS

Movimiento asociativo abierto a la ciudadanía que defiende los derechos de las personas con discapacidad intelectual o del desarrollo y de sus familias

**Fundación
ATENEA**

Fundación especializada no sólo en la intervención con colectivos en situación de vulnerabilidad sino también en investigación social

ASECEDI

Asociación de entidades de Centros de Día profesionalizados dedicadas al tratamiento de las adicciones tanto tóxicas como no tóxicas, en dispositivos tan específicos como son dichos Centros de Día

Dichas entidades colaboradoras, proporcionan a la SGIIPP un extenso catálogo de servicios, que la solución **Predictive Data Justice** ha de incorporar como entrada para su motor de recomendaciones (Next Best Action), con el fin de proponer el subconjunto de servicios óptimo y personalizado para cada tipo de recluso:

- Programas de violencia de género:
 - ✓ Violencia de género: Programa de intervención para agresores.
 - ✓ El delito de violencia de género y los penados extranjeros.
- Programas de drogodependencias:
 - ✓ Campañas de prevención y sensibilización ante el consumo de drogas.
 - ✓ Proyecto Cardenal Martini.
 - ✓ Programa Terapéutico Loyola.
 - ✓ Programa base para consumidores de opiáceos y poli consumidores.
 - ✓ Programa de cocaína.
 - ✓ Tratamiento con Metadona.
- Programas de alcoholismo:
 - ✓ Campañas de prevención y sensibilización ante el consumo de alcohol.
 - ✓ Programa específico para personas con problemas de dependencia alcohólica.
- Programas de ludopatía:
 - ✓ Programas para adicciones sin sustancia: Especialmente ludopatías y otras adicciones comportamentales.

- Programas de integración social y actividades rehabilitadoras para personas reclusas y ex-reclusas con discapacidad intelectual.
- Programas de atención a las personas con necesidades de atención integral socio-sanitaria.
- Programas de atención a las personas con necesidades educativas o de inserción laboral.
- Programas de salud mental.
- Programas para el control de la agresión sexual.
- Programas de la violencia doméstica.
- Programas de la violencia filio parental.
- Programas formativo/laborales:
 - ✓ Programas de alfabetización, español para extranjeros, formación profesional, e inserción laboral.
 - ✓ Talleres de orientación y búsqueda de empleo.
- Programas de la seguridad vial:
 - ✓ Talleres de seguridad vial.
 - ✓ Apoyo o participación en programas de educación vial, primeros auxilios, y socorrismo.
- Programas para mujeres:
 - ✓ Unidad dependiente Ellacuría.
 - ✓ Programas Strada.
- Programas dirigidos a personas con discapacidad:
 - ✓ Apoyo a discapacitados físicos y psíquicos.
 - ✓ Apoyo a personas mayores y dependientes.
 - ✓ Apoyo a enfermos terminales.

- Programas para adolescentes.
- Programas para personas extranjeras.
- Programas de motivación intra-penitenciarios y comunidad terapéutica intra-penitenciaria.

3.3 Contactos y entrevistas

Para poder abordar con garantías el desarrollo de nuestro proyecto, hemos mantenido diversas reuniones con varios representantes que ejercen su labor profesional en el ámbito jurídico-penitenciario.

En primer lugar, quisimos conocer de primera mano la problemática que podría suscitar que nuestra solución utilizase los datos generados durante el proceso jurídico. Para ello, mantuvimos un encuentro con el Jefe de Sección del **Comité Nacional de Estadística Judicial (CNEJ)**, órgano dependiente del **Consejo General del Poder Judicial (CGPJ)**. Su acogida hacia nuestro proyecto fue bastante positiva, ya que actualmente tienen puesto en marcha un proyecto de minería de datos, a partir de las sentencias que se publican en el **Centro de Documentación Judicial (CENDOJ)**, un área de jurisprudencia del CGPJ. Han elaborado un diccionario de datos a través del CENDOJ, con la intención de realizar conteos sobre el vocabulario jurídico utilizado, y así poder elaborar diversas estadísticas y estudios.

Además, nos proporcionaron un portal web donde se publican dichas estadísticas, así como las instituciones con las que colaboran para elaborar ciertos estudios, sobre todo en el ámbito delictivo-penal.

Posteriormente, contactamos con el **Observatorio de Violencia Doméstica**, y mantuvimos un encuentro con la responsable de la **Unidad de Igualdad y Violencia de Género**. En la actualidad, elaboran sus estudios estadísticos anuales con datos extraídos de unas **50 sentencias**, que analizan de una forma completamente manual, por lo que su acogida hacia nuestro proyecto fue bastante positiva y optimista. Además, nos animó a continuar hacia delante con el proyecto, dada su posible viabilidad, utilizando para ello los fondos documentales del CENDOJ proporcionados por el MJ.

Después, mantuvimos un encuentro con profesionales del **Cuerpo Nacional de Policía (CNP)**, para conocer más en detalle el contenido de los **atestados e informes** que se generan cuando cualquier delincuente comete un delito o falta, así como su traslado al correspondiente juzgado, y su posterior paso a la SGIIPP.

Hicimos especial hincapié en conocer con detalle los Sistemas de Información que actualmente utilizan, y su conectividad con los distintos juzgados para compartir dicha información. Bien es cierto que aún queda mucho por mejorar al respecto, aunque en la actualidad existen diversos planes recogidos en los correspondientes presupuestos generales para acometer la interoperabilidad de todos los sistemas jurídicos y policiales.

Con posterioridad, contactamos con la **Subdirección General de Nuevas Tecnologías del Ministerio de Justicia (SGNTMJ)**, con el fin de llegar a conocer los **sistemas de digitalización** que tienen implantados actualmente. Nos presentó el **Expediente Judicial Electrónico (EJE)**, el proyecto tecnológico “estrella” implantado en la anterior legislatura en la Audiencia Nacional (AN) y en otras 10 provincias españolas.

Por último, contactamos con la **Agrupación de los Cuerpos de la Administración de Instituciones Penitenciarias (ACAIP)**, y con la **Fundación Atenea** para conocer las estadísticas publicadas sobre la población reclusa en España en comparación con la media de la Unión Europea, condiciones de los reclusos y empleados públicos de la SGIIPP, planes de reinserción laboral, así como las recomendaciones del **Consejo Europeo**, en cuanto a la utilización de las medidas alternativas a la privación de la libertad en centros penitenciarios, mediante una reeducación y reinserción social eficiente, base que pretende respetar nuestra solución.

Finalmente, hemos podido verificar que las diferentes organizaciones visitadas utilizan unas herramientas analíticas descriptivas, y el volumen de la información analizada para la realización de sus estadísticas es bastante pequeño. Percibimos un **enorme entusiasmo por nuestro proyecto**, ya que podrá mejorar sus estadísticas descriptivas al tratar nuestra solución un mayor volumen de información, realizando un análisis mucho más profundo, utilizando unos modelos analíticos-predictivos novedosos, creados de manera específica para tal fin. Y nuestra conclusión es clara: **la AGE tiene la capacidad y la intención de modernizar sus procesos analíticos, y están muy interesados en contar con nuestra solución.**

3.4 Otros posibles clientes

Nuestro proyecto **Predictive Data Justice** es perfectamente escalable a otros clientes del **panorama internacional**. Se trata de una nueva oportunidad de negocio, que sin embargo cuenta con un pequeño inconveniente en lo referente al lenguaje, ya que nuestra solución soportará únicamente (y de momento) el idioma

español. Por lo tanto, nuestra solución se podrá implantar tal cual, es decir sin necesidad de modificarla, en otros países que compartan ese mismo idioma. Otra opción algo más a futuro consistirá en desarrollar un cambio de idioma en nuestra solución, como fase de expansión a otros países no hispanohablantes.



Figura 3-3. Tasa de población reclusa en Sudamérica

Por ejemplo, en el caso concreto de **Sudamérica**, se puede comprobar que existen países con una tasa de población reclusa bastante similar a la de España.

Del mismo modo, otros posibles interesados en nuestro proyecto, sin ser el objetivo principal del mismo, son los **abogados**, constituidos o no en bufetes. Actualmente, manejan mucha información sobre sentencias, tipos de jueces y juzgados, estadísticas sobre casos con sentencias favorables y desfavorables, etc., lo que francamente supone un ingente esfuerzo el manejar y gestionar toda esa información. Por lo general, la base de su negocio reside en la experiencia y en la excelencia de sus recursos. La idea de enfocar este proyecto hacia los abogados, reside en la oportunidad de proporcionarles una nueva forma de obtener información estadística sobre futuros procesos judiciales, teniendo en cuenta los perfiles de los implicados en los mismos. De esta forma, los abogados podrán predecir el porcentaje de éxito de sus futuros procesos judiciales, con el objetivo de minimizar riesgos y estimar con mayor precisión el éxito de cada caso.

4 ¿Quiénes somos?

4.1 Definición instrumental de la empresa

Predictive Data S.L. es una empresa especializada en el diseño, desarrollo e integración de soluciones Big Data & Business Analytics dentro del **ámbito público internacional** (Administraciones, Gobiernos, Ministerios, etc.), extrayendo automáticamente información de todo tipo de contenidos para darle valor, aplicando tecnologías semánticas (minería de texto) y un potente motor de recomendaciones.

Tenemos distintas soluciones o productos en el mercado internacional para los sectores de Sanidad, Transporte y Servicios de carácter público.

En el ámbito jurídico-penitenciario, se plantean nuevos retos y requerimientos que los actuales sistemas de gestión documental no son capaces de soportar.

Dada la alta relevancia de los contenidos presentes en la **documentación jurídica y penitenciaria**, resulta imprescindible identificar y extraer, de forma automática, determinados conceptos que se manejan, utilizando para ello unos modelos que denominamos como ontologías.

Estos conceptos son automáticamente reconocidos y extraídos de los contenidos documentales, gracias a la solución analítica que se propone en **Predictive Data Justice**.

A través de **Predictive Data Justice**, la asignación de recursos humanos y materiales - optimizando el gasto presupuestario-, las situaciones de prisión provisional y la conveniencia de conceder o no un determinado beneficio penitenciario, estarán mejor apoyados por nuestra solución, que se alimenta de grandes bases de datos y opera sobre potentes y novedosos modelos y algoritmos previamente definidos

4.2 Misión, visión y valores

Predictive Data S.L. destaca su compromiso con la misión, visión y valores para el desarrollo social en el ámbito público, y la implantación de las nuevas tecnologías analítico-predictivas basadas en **Big Data**, para dar un servicio mejorado a la sociedad.

4.2.1 Misión

Nuestra misión consiste principalmente en promover planes alternativos y apoyar las decisiones tomadas por las diferentes instituciones públicas, ofreciendo asistencia en la integración y el desarrollo de soluciones Big Data mediante técnicas analíticas de grandes volúmenes de datos, para la correcta adecuación y mejora de la efectividad en la asignación de los recursos públicos.

4.2.2 Visión

Nuestra visión consiste fundamentalmente en ser un referente, reconocido tanto a nivel nacional como internacional por su gran contribución en servicios públicos, desarrollando proyectos y servicios innovadores de alta calidad, así como rentables a nivel económico y social.

4.2.3 Valores

Tabla 4-1. Valores de Predictive Data

Organización multidisciplinar, con amplia diversidad de perfiles profesionales

Compromiso con el cliente

Respeto al código de conducta

Integridad y ética profesional

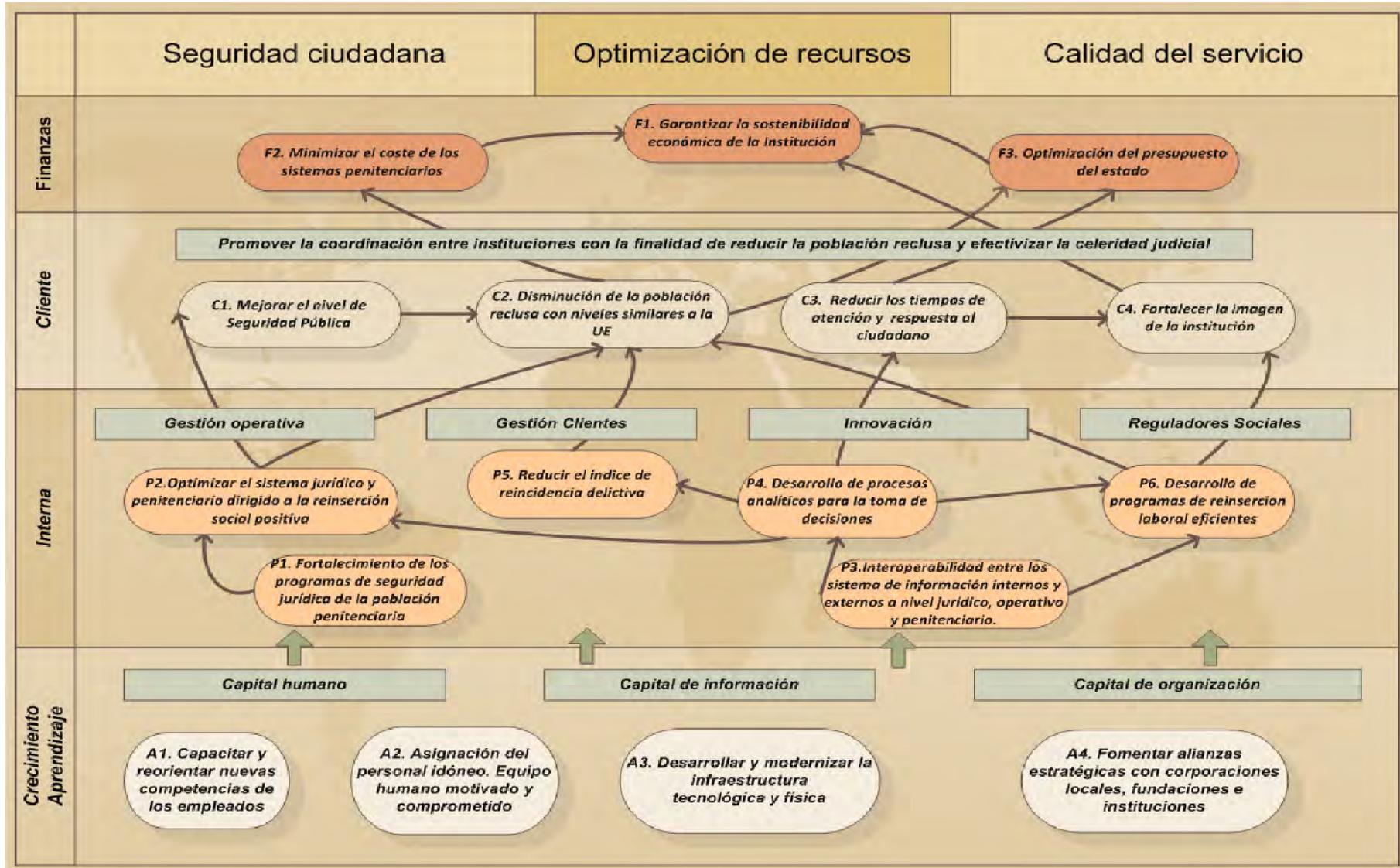
Experiencia y Know-how: Cliente, tecnología y Big Data

4.3 Plan estratégico

La estrategia empresarial de Predictive Data, se traduce en el Mapa estratégico, bajo las 4 perspectivas del modelo de Kaplan y Norton.

4.3.1 Mapa estratégico

Para ejecutar el plan estratégico se ha elaborado el siguiente plan estratégico con los siguientes objetivos estratégicos.



4.3.2 Indicadores clave (KPIs)

Para la consecución y control de todos los objetivos diseñados para nuestro mapa estratégico, hemos elaborado los siguientes indicadores clave (KPIs) con sus respectivas metas y periodo temporal.

Tabla 4-2. Indicadores clave de desempeño (KPIs)

Perspectiva	Indicador	Meta	Tiempo
Financiera	Carga financiera sobre presupuestos liquidados	11,3 *	2-3 años
	Costes fijos / Costes variables	0,6	3 años
	Coste medio por preso al día.	1650	3 años
	Ahorro de las partidas presupuestarias	70M	3 años
	Superávit (o Déficit) por habitante: Resultado presupuestario ajustado / Número de habitantes	0,9 % PIB	3 - 4 años
Cliente	Porcentaje de delitos penales	< 3,5%	2 años
	Porcentaje de reincidencia. Tasa de reincidencia	38%	4 años
	Tasa de población reclusa/100.000 habitantes	130	4 años
	% población reclusa extranjera	25%	2 años
	Tiempo medio en dictaminar sentencias penales firmes	150 días	3 años
	Tiempo medio de un procedimiento judicial	255 días (< 25 días) **	2 -3 años

	Porcentaje de satisfacción de la población ciudadana	6 (0-10)** *	3 años
	Nivel de Confianza en las instituciones públicas en el ámbito jurídico	60%	3 años

* El límite impuesto en las CCAA es de 25,00

** Respecto a la actualidad

*** Encuestas satisfacción (0-10 puntos)

Perspectiva	Indicador	Meta	Tiempo
Procesos	Tasa de condena media	14 meses (<5)**	2-3 años
	Nª de actividades alternativas a la pena impuesta	> 35% actual	2-3 años
	Porcentaje de reinserciones sociales satisfactorias	75%	2-3 años
	Porcentaje de delitos cometidos tras realizar programas de reinserción	< 30 %	3 años
	Número de Sistemas digitalizados	>70%	1-2 años
	Porcentaje de Sistemas Interconectadas	90%	1-2 años
	Tiempo medio de respuesta en atender peticiones internas entre instituciones	< 3 días	1-2 años
	Porcentaje de internos penitenciarios clasificados	95%	3 años
	Porcentajes de algoritmos matemáticos implementados	95%	3 años
	Tasa de usabilidad: Número de programas predictivos usados/ Numero de programas instalados	90%	1,5 años
	Tasa de reincidencia delictiva	38%	4 años
	Edad media de presos reincidentes	35 años	2-3 años
	Porcentaje de iniciativas y/o programas exitosos	70%	1,5- 2 años

Crecimiento - Aprendizaje	Ratio de Eficiencia: Número de programas realizados / Número de programas previstos	85%	2 años
	Nº de horas de capacitación por empleado	120 horas	1-2 años
	Porcentaje de asistencia completa por cursos	95%	1-2 años
	Índice de satisfacción personal	7.5 p (0-10) ^{***}	1-2 años
	Personal que supera evaluación del desempeño	85%	1-2 años
	Convenios interinstitucional y/o multisectorial firmados	6	2-3 años

4.3.3 Debilidades, amenazas, fortalezas y oportunidades

El análisis **DAFO**, es una metodología de estudio de la situación de nuestro proyecto, analizando sus características internas (**debilidades y fortalezas**) y su



situación externa (**amenazas y oportunidades**) en una matriz cuadrada.

Figura 4-1. DAFO

4.3.3.1 Fortalezas

- ✓ Alto conocimiento tecnológico de nuestro equipo gracias a nuestra experiencia con las nuevas tecnologías, modelos predictivos y motor de recomendaciones.
- ✓ Prueba de concepto validada con alguno de nuestros colaboradores. Gracias a las entrevistas mantenidas, el equipo ha incrementado su conocimiento sobre la materia para enfocar, plantear y aterrizar el

problema de negocio y su solución. Ver el **Capítulo 3.3 Contactos y entrevistas**.

- ✓ Avalado por estudios estadísticos. Todos los estudios estadísticos existentes inciden en la necesidad de conocer bien en detalle los factores psicosociales para prevenir la reincidencia. Nuestro equipo se encuentra perfectamente formado en la materia, conocedores en todo momento de los numerosos estudios existentes y que avalan la necesidad de incidir sobre esos factores para reducir la reincidencia delictiva.

4.3.3.2 Debilidades

- ✓ No consecución de los objetivos cuantitativos marcados al inicio del proyecto por infraestimación del coste total del proyecto.
- ✓ No existe un profundo conocimiento judicial y legal de los procesos operacionales del cliente.

4.3.3.3 Oportunidades

- ✓ Proceso judicial digitalizado.
- ✓ Utilización del Expediente Judicial Electrónico, con el objetivo de complementar a la información como propuesta de valor del proyecto.
- ✓ Disponemos de un gran volumen de datos para analizar y sacar conclusiones. Es una oportunidad para aplicar las tecnologías de Big Data, junto con el procesamiento en paralelo, los modelos analítico-predictivos y un potente motor de recomendaciones, para proporcionar conocimiento y su puesta en valor.
- ✓ Fuentes de información en Texto proveniente de sentencias, autos, atestados policiales, informes, etc.
- ✓ No se trata toda la información que se encuentra disponible. Actualmente, el entorno jurídico-penitenciario debe almacenar por ley toda la información procesal. Predictive Data obtendrá ventaja de su explotación para dar un servicio de alta calidad.
- ✓ Actualmente, en el ámbito jurídico-penitenciario no se emplean ni modelos analítico-predictivos ni un motor de recomendaciones, que sirvan como ayuda para mejorar en la toma de decisiones.

4.3.3.4 Amenazas

- ✓ Incertidumbre en el mantenimiento de los presupuestos.
- ✓ Cambios de organizativos y políticos.
- ✓ Déficit presupuestario.
- ✓ En general, las amenazas externas vienen de la mano de la estabilidad en la organización del estado y su presupuesto.

4.4 Modelo de negocio

Hasta ahora, hemos comentado en detalle cuál es nuestro problema de negocio, dando a conocer algunas de las estadísticas existentes en la actualidad sobre la delincuencia, el proceso judicial, la reinserción social, la tasa de población reclusa, etc. (ver el Capítulo 2.2, Comparativa de España frente a Europa). A su vez, hemos definido cuáles son las entidades colaboradoras, organizaciones, instituciones, fundaciones, etc. que intervienen en alguna fase del proceso jurídico-penitenciario (ver el Capítulo 3 ¿Quiénes son nuestros clientes?), y que por ello se convierten en potenciales clientes de nuestra solución. Por lo tanto, estamos en disposición de presentar ya nuestro modelo de negocio, que representamos mediante el siguiente un cuadro (o diagrama Canvas).



Figura 4-2. Modelo de negocio

El modelo de negocio representa las siguientes relaciones entre los intervinientes del negocio (colaboradores, socios y clientes), la propuesta de valor, canales de comunicación y los usuarios finales.

Los **socios clave** del proyecto se encuentran enmarcados en las **entidades colaboradoras** del Ministerio de Justicia y del Ministerio del Interior, como pueden ser la Fiscalía General del Estado, las Fuerzas y Cuerpos de Seguridad del Estado (Cuerpo Nacional de Policía y Guardia Civil), la Secretaría General de Instituciones Penitenciarias, así como otras entidades que nutren de estudios estadísticos al sistema ya sean bien internas o externas, como pueden ser el Consejo Superior del Poder Judicial o Eurostat.

Del mismo modo, para articular el proyecto es necesario contar con **elementos clave tecnológicos**, como pueden ser los modelos analítico-predictivos y el desarrollo actual de las herramientas de minería de datos, así como el aterrizaje de nuevos conceptos, como por ejemplo puede ser el del ciclo reincidente-disidente, referido a la reiteración en la comisión de delitos por parte de un individuo. Los recursos tecnológicos vienen de la mano de las **nuevas tecnologías de Big Data**, plataformas de procesamiento que detallamos a continuación en el Capítulo 5 Solución.

Nuestra **propuesta de valor** es un sistema de recomendación sobre la conveniencia de conceder beneficios penitenciarios y una revisión del código penal en cuanto al endurecimiento de las penas, midiendo, a través de un modelo predictivo, el riesgo que los reclusos tienen de volver a cometer reincidencia delictiva.

Aparte de las instituciones públicas comprometidas con el sistema jurídico-penitenciario, consideramos también como **cliente final al conjunto de la sociedad, incluyendo principalmente en este a los reclusos, ya que con su reeducación y reinserción social, se obtiene una seguridad ciudadana mejorada.**

Lo más importante que cabe destacar de este proyecto es su **explícito y fuerte valor social**, que garantiza un servicio público en el ámbito de la Justicia y la seguridad ciudadana eficientes y de gran calidad, tal y como exigen tanto la ciudadanía como la población penitenciaria a través de sus derechos reconocidos, beneficiándose de todo ello el conjunto de la sociedad.

5 Solución

5.1 Definición del producto

Predictive Data Justice es el producto que materializa la propuesta de valor descrita en esta memoria. Tiene la capacidad de:

- ✓ Conectores con datos estructurados y no estructurados.
- ✓ Almacén de información (Data Lake) altamente escalable.
- ✓ Ejecución de un modelo predictivo de riesgo de reincidencia avalado por nuestra experiencia.
- ✓ Un motor de recomendaciones que ayuda en la toma de decisiones.
- ✓ Una aplicación visual, deslocalizada basada en WEB, versátil, intuitiva y customizada al usuario final.

Predictive Data Justice captura datos internos y externos, estructurados y no estructurados, realiza su tratamiento con **modelos matemáticos analítico-predictivos** previamente contruidos y entrenados a partir de los **datos socio demográficos, penales, penitenciarios, familiares, escolares, laborales y de salud mental** de los reclusos, para finalmente, recomendar de una forma

específica a cada recluso que presente un elevado riesgo de cometer reincidencia delictiva, **posibles tratamientos**, dentro del catálogo de servicios proporcionados por las entidades que colaboran con la institución, para lograr una completa rehabilitación del recluso.



Figura 5-1. Logotipo del producto

El novedoso paradigma de **Big Data** aparece gracias a la utilización de soluciones software “Open Source”, como por ejemplo el ecosistema del proyecto **Apache Spark**, un motor para el procesamiento general y rápido de datos a gran escala, realizado por la fundación americana Apache Software Foundation (ASF).

Para el almacenamiento y procesamiento de datos no estructurados provenientes fundamentalmente de redes sociales e Internet, así como de documentos previamente digitalizados (atestados y denuncias policiales, sentencias y autos judiciales, diligencias, y demás informes etc.), el producto también se apoya en herramientas “Open Source”, que se encuentran disponibles en el ecosistema Apache Spark.

Tratamos siempre de buscar el punto óptimo de equilibrio entre el coste de recursos especializados en herramientas “Open Source”, el coste equivalente de las licencias de paquetes software de utilización sencilla y el soporte de los fabricantes incorporados.

5.2 Análisis de las fuentes de datos

Para alimentar la solución propuesta y adaptar el producto a la solución específica del cliente se propone la utilización de las siguientes fuentes de datos:

- **Expediente Judicial:** Sentencias, atestados y autos obtenidos del CENDOJ y el EJE (Expediente Judicial Electrónico).

- **Expediente Penitenciario:** Fichero de datos de carácter personal del Sistema Informático Social Penitenciario (SISPE) y los informes generados por los educadores, psicólogos, sociólogos, pedagogos y trabajadores sociales del centro.
- **Redes sociales e Internet:** Contenido publicado en Internet que alimenta el motor de recomendaciones para evaluar la efectividad de las acciones.

5.2.1 Expediente Judicial

Predictive Data Justice es capaz de obtener la información relevante de todo el proceso judicial del recluso utilizando minería de texto sobre la documentación existente en los sistemas como por ejemplo el CENDOJ o el EJE.

El Centro de **Documentación Judicial (CENDOJ)** es el **órgano técnico del Consejo General del Poder Judicial (CGPJ)** que se ocupa de la publicación oficial de la jurisprudencia (es el conjunto de sentencias dictadas por los Juzgados y Tribunales, y la concreta doctrina o criterio que las mismas establecen) (art. 560.1.10ª de la L.O. 4/2013), así como de las demás competencias en el ámbito de la documentación y de los servicios de gestión del conocimiento.

En ejercicio de esta misión, el **CENDOJ** ofrece servicios de apoyo e información a los miembros de la Carrera Judicial. Para ello facilita el acceso a todo tipo de fuentes documentales (jurisprudencia, legislación y publicaciones) que son utilizados en el desarrollo de la actividad judicial, presta servicios de búsqueda de documentación y se ocupa tanto de la red de bibliotecas judiciales como del desarrollo de los entornos web del Consejo General del Poder Judicial, Tribunal Supremo, Audiencia Nacional y Tribunales Superiores de Justicia.

En su labor de difusión de la jurisprudencia española, el **CENDOJ** pone a disposición del público en general, de forma gratuita y a través de la página web del Poder Judicial, las resoluciones judiciales dictadas por los Tribunales españoles.

Aproximadamente, en la actualidad hay cerca de seis millones de autos y sentencias publicadas en el CENDOJ en formato digital (.pdf).

Del mismo modo, la información Judicial puede ser extraída de otras fuentes de datos como el **Expediente Judicial Electrónico (EJE)**.

5.2.2 Expediente Penitenciario

El **Sistema Informático Social Penitenciario (SISPE)** es un gran banco de datos informático que empezó a funcionar en el año 2008, cuando según el BOE nº 245 del Viernes 10 de Octubre de 2008, salió publicada una orden correspondiente del Ministro del Interior, por aquel entonces D. Alfredo Pérez Rubalcaba, que formaba parte del Plan AVANZ@ del Gobierno de España para el “desarrollo de la sociedad de la información”.

Contiene informaciones íntimas y confidenciales sobre la totalidad de personas presas en las cárceles dependientes del Ministerio del Interior. Conectando con él, cualquier funcionario, policía, juez, tribunal o fiscal e incluso trabajadores y educadores sociales podrán conocer, no sólo los antecedentes del detenido o preso, sino todo su historial personal incluido en su expediente penitenciario: no sólo los datos habituales sobre condena, reincidencia, sanciones disciplinarias o “tratamiento penitenciario”, sino también todo tipo de información confidencial, como “tests” psicológicos o cualquier otro material elaborado por los equipos de observación, información sobre su ambiente familiar, “situación de su entorno desde que fue encarcelado o información sobre su vivienda habitual y si residía en un ambiente marginal, urbano o rural”, sobre el nivel educativo y los cursos realizados en prisión, así como todos los detalles relacionados con su vida laboral; información tan íntima como las enfermedades contraídas (con especial atención al SIDA), salud mental, discapacidades, tratamientos médicos recibidos o personas con las que se relaciona habitualmente, y también información detallada sobre consumo de drogas.

Por tanto, el fichero **SISPE** permite disponer de información relativa a múltiples aspectos considerados como factores asociados a la reincidencia delictiva y al comportamiento delictivo violento. Se divide principalmente en las siguientes categorías:

- Datos socio demográficos.
- Datos penales.
- Datos penitenciarios.
- Datos familiares.
- Datos escolares y laborales.
- Datos del nivel socioeconómico.

- Datos de los grupos de relación, ocio y tiempo libre.
- Datos de conductas adictivas.
- Datos de salud mental.

Además, se incluyen también sendas escalas psicológicas para la evaluación de la psicopatía y de la personalidad límite de los reclusos.

Tabla 5-1. Estructura y tipos de datos contenidos en el fichero SISPE.

Nombre	Comentario	Tipo de dato
DSOC_CODIGO	Identificador del objeto	NUMBER(10)
DSOC_CODEXP	Expresión de identificador	NUMBER(10)
ESTU_CODIGO	Código del nivel de instrucción o estudios	VARCHAR2(2)
DSOC_FREG	Fecha en la que se va a guardar la información. Para el histórico	DATE
DSOC_DINFAM	Situación Familiar – Dinámica Familiar	VARCHAR2(512)
DSOC_INSTIT	Situación Ambiental – Si el individuo está institucionalizado o no	NUMBER(1)
DSOC_CINST	Situación ambiental – Tipo de centro en el que está institucionalizado en el caso de que así haya sido	CHAR(1)
DSOC_FINICP	Situación ambiental – Fecha desde la que está en Prisión	DATE
DSOC_OBSSA	Situación Ambiental. Otros datos o comentarios del trabajador social al respecto	VARCHAR2(1024)
DSOC_TIPVIV	Situación Ambiental – Datos de la vivienda – Tipo de vivienda	CHAR(1)
DSOC_REGVIV	Situación Ambiental – Datos vivienda – Régimen de la vivienda	CHAR(1)
DSOC_CONVIV	Situación Ambiental – Datos vivienda – Condiciones en las que se encuentra la vivienda	VARCHAR2(1024)

DSOC_BARRIO	Situación Ambiental – Datos vivienda – Entorno Social – Tipo de barriada en la que vive el individuo	CHAR(1)
DSOC_NUCLEO	Situación Ambiental – Datos vivienda – Entorno Social – Núcleo de población en el que vive	VARCHAR2(1024)
DSOC_GRUPRL	Situación Ambiental – Datos vivienda – Entorno Social – Grupo de relaciones	VARCHAR2(1024)
DSOC_NIVEST	Situación Educativa – Nivel de estudios alcanzado	VARCHAR2(2)
DSOC_ESTUD	Situación Educativa – Especificación de los estudios alcanzados por el individuo en el caso de tener estudios universitarios	VARCHAR2(128)
DSOC_EDADDE	Situación Educativa – Edad en la que dejó los estudios	NUMBER(3)
DSOC_MOTDE	Situación Educativa – Motivo por los que dejó los estudios	VARCHAR2(512)
DSOC_OBSEST	Situación Educativa – Comentario sobre la formación por parte del trabajador social (Observaciones)	VARCHAR2(1024)
DSOC_SITLAB	Situación Laboral – Situación laboral al ingreso	CHAR(1)
DSOC_OBSLAB	Situación Laboral – Otros comentarios u observaciones por parte del trabajador social	VARCHAR2(1024)
DSOC_TJTSAN	Situación Salud – Número de la tarjeta sanitaria en el caso de que la tenga	VARCHAR2(20)
DSOC_CSALUD	Situación Salud – Centro de salud que le corresponde en el momento en el que esté en libertad	VARCHAR2(128)
DSOC_DISCAP	Situación Salud – Tipo de discapacidad que puede presentar el individuo	VARCHAR2(50)
DSOC_PORDIS	Situación Salud – Porcentaje de minusvalía	NUMBER
DSOC_MINUSV	Situación Salud – Posibilidad de	NUMBER(1)

	que tenga minusvalía o no	
DSOC_SIDA	Situación de Salud – Si es seropositivo	NUMBER(1)
DSOC_OBSSAN	Situación Salud – Comentario u observaciones sobre la situación de salud que presenta el individuo por parte del trabajador social	VARCHAR2(1024)
DSOC_DROGAS	Situación Salud – Consumo de drogas –drogas consumidas	VARCHAR2(4)
DSOC_EDADID	Situación Salud – Consumo de drogas –Edad de inicio en el consumo de drogas	NUMBER(3)
DSOC_EDADFD	Situación Salud – Consumo de drogas –Edad de finalización en el consumo de drogas	NUMBER(3)
DSOC_TTO	Situación Salud – Consumo de drogas – Tratamiento recibido para rehabilitación de un problema de drogadicción	VARCHAR2(512)
DSOC_LUGTTO	Situación Salud – Consumo de drogas – Lugar donde ha recibido el tratamiento de rehabilitación	VARCHAR2(50)
DSOC_FITTO	Situación Salud – Consumo de drogas – Fecha de Inicio del tratamiento	DATE
DSOC_FFTTO	Situación Salud – Consumo de drogas – Fecha de finalización del tratamiento recibido	DATE
DSOC_OBSDRO	Situación Salud – Consumo de drogas – Comentario u observaciones del trabajador social sobre la situación, del consumo de drogas por parte del individuo	VARCHAR2(1024)
DSOC_SITECO	Situación Económica – Descripción	VARCHAR2(1024)
DSOC_INTERP	Intervención Social – Interpretación Diagnóstica Social del trabajador social tras la entrevista realizada	VARCHAR2(1024)
DSOC_PROBS	Intervención Social – Problemas a tratar	VARCHAR2(1024)

DSOC_MOTIVS	Intervención Social – Motivaciones y compromisos por parte del individuo	VARCHAR2(1024)
DSOC_OBJTVS	Intervención Social – Objetivos	VARCHAR2(1024)
DSOC_RECURS	Intervención Social – Recursos	VARCHAR2(1024)
DSOC_OBSSOC	Intervención Social – Comentarios u observaciones del trabajador social	VARCHAR2(1024)

Predictive Data Justice no solo obtiene la información de SISPE sino también es capaz de realizar **minería de texto** sobre todos los **informes** que en prisión generan los actores que se encuentran involucrados en el proceso de reinserción del individuo como son los educadores, trabajadores sociales, psicólogos, sociólogos y pedagogos.

5.2.3 Contenido publicado en Internet y redes sociales

Predictive Data Justice está preparado para realizar un rastreo en Internet en busca de información relevante a "reputación" relacionada con las acciones registradas en el motor de recomendaciones. Esta fuente de datos es primordial para evaluar el **impacto social** de las acciones propuestas por el motor de recomendaciones con el fin de realizar un proceso de mejora continua en la toma de decisiones.

5.3 Modelo lógico

A lo largo del presente capítulo, se describe el modelo lógico diseñado para nuestra propuesta de solución tecnológica.

Partiendo de las fuentes de datos estructuradas y no estructuradas (explicadas en el capítulo 5.2 Análisis de las fuentes de datos), se construye el **Data Lake** sobre el cual se asienta la solución **Predictive Data Justice**. El acceso a dicho Data Lake estará securizado, mediante un módulo de gestión de usuarios capaz de controlar los accesos, de tal forma que únicamente dispondrán de ello determinados perfiles de usuario, como por ejemplo el administrador del sistema, y ciertos usuarios a los que se den permisos muy específicos.

Todas las posibles acciones que puedan realizar los usuarios, quedarán registradas en ficheros "log" que serán almacenados convenientemente en el Data Lake.

Para evaluar el riesgo de reincidencia delictiva de un determinado recluso, el usuario del sistema mandará una determinada petición, que tras la ejecución de

un modelo matemático analítico-predictivo, y combinado el resultado obtenido con la acción de un motor de recomendaciones, podrá visualizar el correspondiente riesgo de reincidencia delictiva, junto con un listado de acciones propuestas recomendadas para aplicar sobre el recluso, con el fin de lograr reducir el riesgo de reincidencia delictiva del recluso. Todas esas acciones se introducen en el motor de recomendaciones a través de un gestor, que cuenta con unos permisos muy específicos para ello.

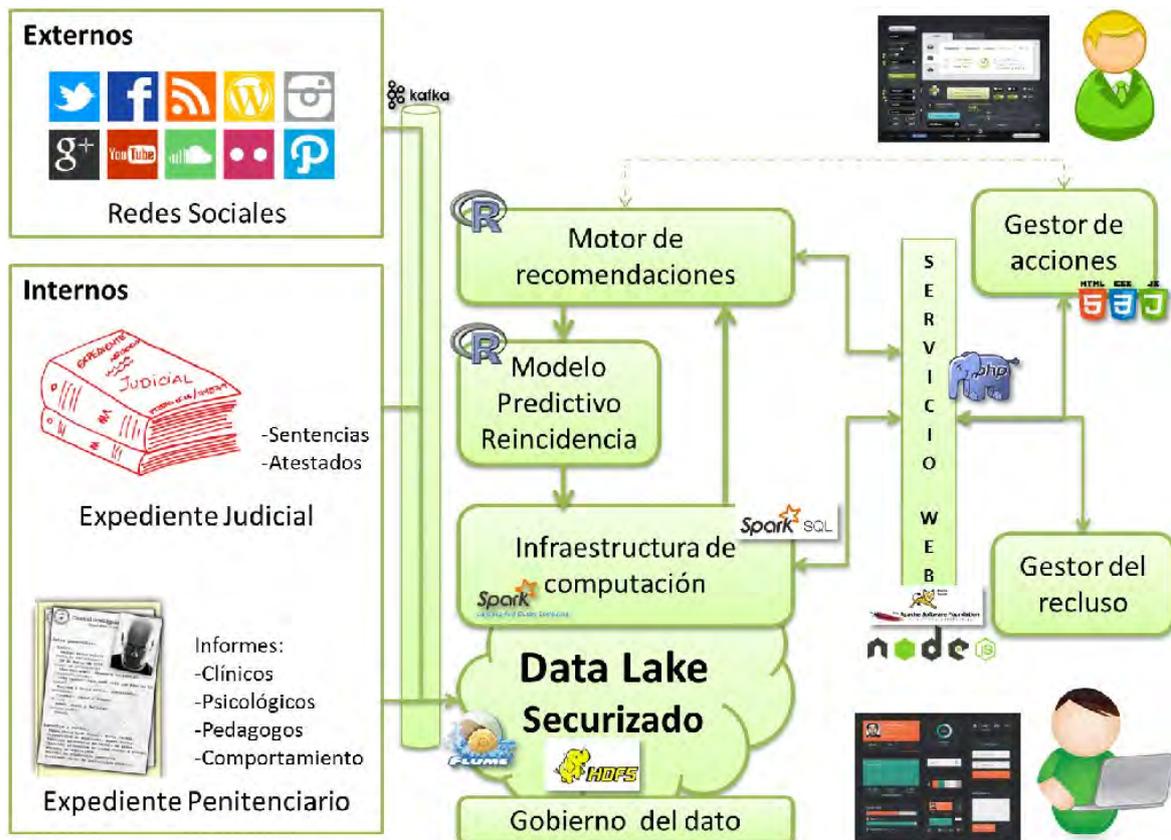


Figura 5-2. Modelo lógico

5.4 Arquitectura

Desde el punto de vista de la arquitectura **Predictive Data Justice** utiliza una pila de tecnología "Open Source", una vez combinadas todas ellas entre sí, son capaces de brindar una alta capacidad de rendimiento y potencia de procesamiento, con una gran escalabilidad, flexibilidad, y modularidad semejantes a las herramientas comercializadas bajo licenciamiento.

La siguiente figura representa gráficamente como queda constituida físicamente la arquitectura de la solución.

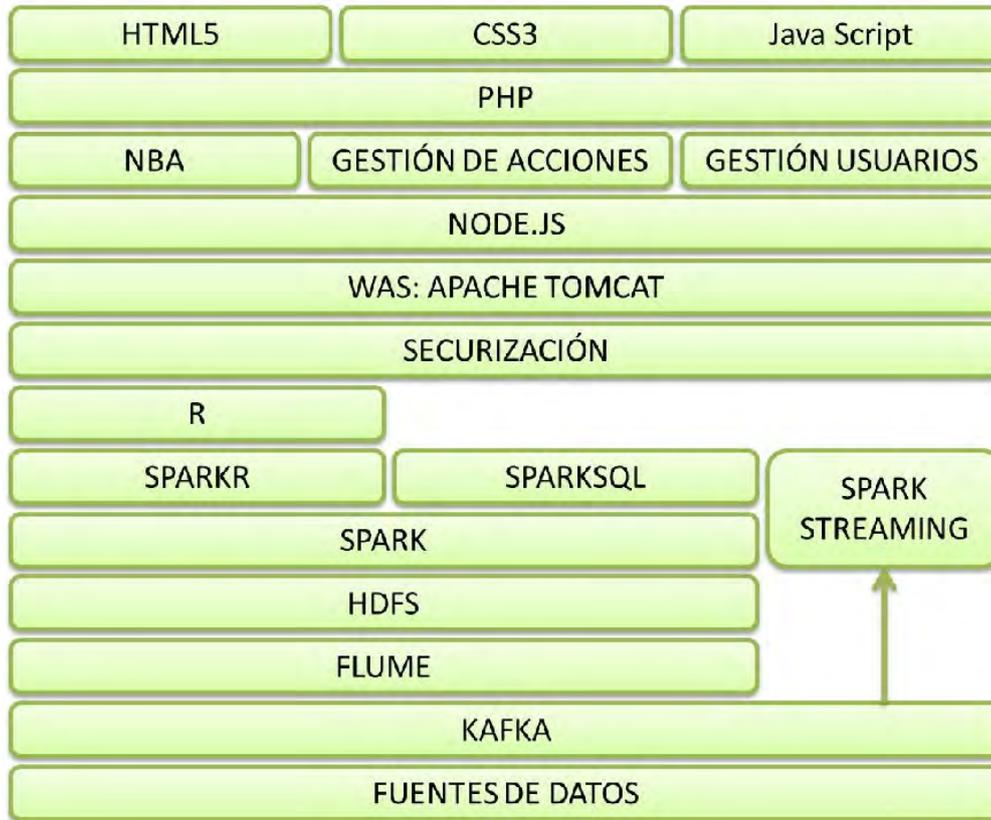


Figura 5-3. Arquitectura física

Leyendo la arquitectura de abajo arriba, la ingesta de los datos (tanto estructurados como no estructurados) se realiza mediante las herramientas Kafka y Flume.

Kafka se utiliza para introducir y desencadenar nuevos eventos dentro de la solución, como por ejemplo puede ser el alta de un nuevo recluso o alertas de cambio de estado producidas por fuentes de datos externas. Además, es un gran gestor de colas, que ingesta datos provenientes de diversas fuentes hacia varios destinos.

Flume se utiliza fundamentalmente para la ingesta en el sistema de datos que provienen de las fuentes de datos previamente establecidas.

Todos los datos relevantes de la solución, quedan almacenados en un Data Lake securizado, tal y como se estableció en el capítulo 5.3 Modelo lógico. En concreto, está compuesto por un sistema de ficheros denominado **HDFS** al que se accede mediante la plataforma **Spark**.

La utilización de la plataforma Spark viene motivada por las siguientes características:

- ✓ Se trata de una tecnología del tipo “Open Source”.
- ✓ Comparada con otras, se trata de una plataforma altamente novedosa e innovadora.
- ✓ Soporta varios lenguajes de programación, como por ejemplo Java, Python y Scala.
- ✓ Incluye diversas soluciones para streaming, bases de datos estructuradas (tipo SQL), ejecución de R, Machine Learning, etc.
- ✓ Mayor rapidez y potencia de cálculo con respecto a otras plataformas de la competencia, como por ejemplo Hadoop.

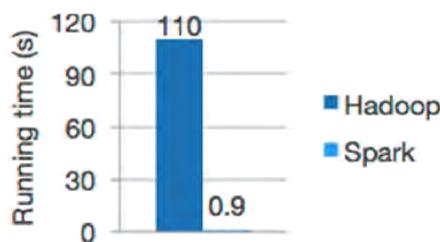


Figura 5-4. Tiempo de regresión logística entre Hadoop y Spark

El acceso para cada uno de los usuarios finales de la solución que se propone en el proyecto **Predictive Data Justice**, se realiza mediante un portal Web construido también íntegramente con herramientas del tipo “Open Source”.

El propósito final de dicho portal Web, consiste en proporcionar unas potentes herramientas de Inteligencia de Negocio para los distintos usuarios de la solución propuesta.

Predictive Data Justice brinda una solución completa y cerrada, que incluye un periodo de mantenimiento para posibles evolutivos.

Los modelos predictivos de reincidencia así como el modelo del motor de recomendaciones son definidos en lenguaje R y ejecutados sobre SparkR.

Los resultados de consultas, logs así como los catálogos estructurados relativos a los reclusos son almacenados igualmente en el Data Lake mediante SparkSQL que ofrece un acceso rápido vía WEB.

El producto incorpora por defecto módulos de escalabilidad de la solución como por ejemplo Node.js y la propia plataforma Spark. El cliente no tiene que preocuparse por cómo va a crecer la aplicación en datos o en usuarios, **Predictive Data Justice** lo soportará a nivel de software.

En la siguiente figura se representan todas las tecnologías utilizadas por **Predictive Data Justice**, desde abajo arriba, desde la ingesta de la información, pasando por el procesamiento, hasta la parte visual del producto.



Figura 5-5. Stack tecnológico de la arquitectura física

Más concretamente, para desarrollar toda la parte del portal Web, se utilizan las siguientes herramientas tecnológicas:

- **HTML5:** Como lenguaje de maquetado utilizado para desarrollar la parte cliente del portal Web.
- **CSS3:** Son las hojas de estilo en cascada utilizadas para también en el desarrollo de la parte cliente del portal Web.

- **Java Script:** Como lenguaje de programación que se incrusta y ejecuta en la parte cliente del portal Web.
- **Apache Tomcat:** Como contenedor y despliegue de toda la parte servidora del portal Web.
- **Node.js:** es un entorno de programación en la capa del servidor (pero no limitándose a ello) asíncrono, con I/O de datos en una arquitectura orientada a eventos y basado en el motor V8 de Google. Fue creado con el enfoque de ser útil en la creación de programas de red altamente escalables, como nuestra propuesta de solución.

5.5 Funcionalidad de la aplicación y perfiles de usuarios

5.5.1 Perfiles de usuarios

Predictive Data Justice incorpora por defecto un módulo de gestión de perfiles de usuario adaptado a las instituciones penitenciarias.

La línea general de atribución de perfiles para los usuarios de la solución se basa en la precisión de las características del usuario, sistematizada a través de la precisión de su perfil o conjunto de funciones profesionales por persona para el caso de los puestos unipersonales o grupo para el resto de puestos.

Mediante los siguientes factores, se determinan la forma en la que establecer los distintos perfiles:

- Localización geográfica (Centro Penitenciario o Servicio Social Externo).
- Funciones o actividad principal a realizar.
- Recursos de información requeridos.
- Procedimientos de actuación.

Por lo tanto, consideraremos todos los siguientes perfiles de usuario con diferentes formas acceso al portal Web de la solución que proponemos en nuestro proyecto **Predictive Data Justice**:

- **Educadores:** Les corresponde la realización de las tareas en materia de observación y tratamiento, de creación de informes, propuestas, coordinación y realización de programas de intervención.

- **Pedagogos:** Les corresponde estudiar al interno desde el punto de vista de su historial escolar, grado cultural y nivel de instrucción, enjuiciando el alcance de sus conocimientos, especialmente los instrumentales, actividades expresivas y aficiones, aportando la información correspondiente al estudio de su personalidad.
- **Sociólogos:** Les corresponde informar de los estudios de personalidad de los reclusos y participar en la ejecución de los tratamientos programados, asistiendo a las reuniones de los mismos.
- **Psicólogos:** Les corresponde estudiar la personalidad de los reclusos desde la perspectiva de la ciencia de la psicología y conforme a sus métodos, calificando y evaluando sus rasgos temperamentales, aptitudes, actitudes y sistema dinámico-motivacional y, en general, todos los sectores y rasgos de la personalidad que juzgue de interés para la interpretación y comprensión del modo de ser y de actuar del observado.
- **Titulados medios** en actividades específicas: Resto de personal compuesto por trabajadores y coordinadores sociales, monitores, etc.

Cada uno de los anteriores perfiles de usuario, podrán introducir nueva información o consultar la ya existente para los reclusos. La figura del educador se encarga fundamentalmente de asignar las actividades de reinserción y de reeducación social a los reclusos, con el aval del comité del centro y dentro de los presupuestos del mismo.

5.5.2 Funcionalidad de la aplicación

A continuación, se enumeran algunas de las funcionalidades más importantes de **Predictive Data Justice**:

- ✓ Obtención de un scoring de riesgos de reincidencia delictiva.
- ✓ Creación de un set de actividades y/o acciones recomendadas que se aplicarán sobre los reclusos en riesgo de cometer reincidencia delictiva.
- ✓ Insertar y/o modificar la información de cada uno de los reclusos.
- ✓ Gestión de distintos perfiles de usuario.
- ✓ Evaluación de cada una de las actividades propuestas por el motor de recomendaciones, con el fin de mejorar la toma de decisiones a la hora de asignarlas a los reclusos.

- ✓ Realización de un seguimiento a posteriori sobre el grado de éxito y/o fracaso de las actividades realizadas por los reclusos.
- ✓ Escaneo en redes sociales obteniendo un ranking de reputación de las acciones propuestas por la penitenciaría.

5.6 Modelo predictivo

En este capítulo se presenta el modelo predictivo de **Predictive Data Justice** y las fases para su creación. Finalmente se muestra el funcionamiento interno del motor de recomendaciones incluido en el producto.

El modelo predictivo de la reincidencia delictiva en población penitenciaria española utilizado por Predictive Data Justice está basado en estudios realizados con un grupo de control.

Este modelo se plantea como un estudio longitudinal con el fin de realizar un seguimiento de los sujetos para evaluar de forma prospectiva la tasa de reincidencia y la incidencia de comportamientos violentos. Los resultados que presentamos en esta memoria proceden del análisis de la información de entrada a la cohorte estadística.

En este estudio afrontamos la reincidencia, considerando reincidente a aquel individuo que ha cometido otros delitos en el pasado siendo su número de ingresos en prisión igual a dos o más.

El eje de entrada del modelo es el **Inventario de Factores de Riesgo e Intervención en Prisiones (IFRIP)** que se presenta en este capítulo y que **Predictive Data Justice** utiliza como base del cómputo de la reincidencia por recluso.

El presente estudio se fundamenta y completa mediante la correcta consecución de las siguientes fases:

- **Fase 1:** Análisis descriptivo de las características de la muestra
- **Fase 2:** Relación entre el análisis de la reincidencia delictiva y otros grupos de variables
- **Fase 3:** Análisis del Inventario de Factores de Riesgo e Intervención en Prisiones (IFRIP)

- **Fase 4:** Modelos explicativos de la reincidencia delictiva

5.6.1 Inventario de Factores de Riesgo e Intervención en Prisiones (IFRIP)

Mediante este instrumento, se pretenden evaluar factores de riesgo para la reincidencia. Está inspirado en el **Level of Service Inventory Revised (LSI-R)**, que es una herramienta de riesgo/necesidad diseñada para estimar los niveles de riesgo de reincidencia en la toma de decisiones, de acuerdo a los requerimientos de supervisión necesarios para los permisos de salida y libertad condicional.

Para la construcción del IFRIP, fue necesaria una adaptación en varios ítems y de los clústeres de riesgo, dado que las circunstancias de utilización son diferentes (reclusos cumpliendo pena de prisión), a las recomendadas en el instrumento original (libertad condicional).

Se compone de 56 ítems que se contestan de forma dicotómica (Sí/No), o bien según una escala de 4 puntos (0, 1, 2 y 3) en función de si el ítem se aplica más o menos al recluso evaluado. Los ítems se agrupan formando 10 clústeres que conforman indicadores de riesgo de reincidencia englobando factores de riesgo estáticos (inmodificables, como pueden ser abusos sexuales y/o malos tratos sufridos en la infancia, historia de violencia anterior, quebrantamiento de órdenes de alejamiento, etc.) y factores de riesgo dinámicos (modificables, como pueden ser el consumo de alcohol, celos, dificultades económicas, actitudes machistas, trastornos afectivos, etc.) o necesidades criminógenas.

Los factores de riesgo considerados en el IFRIP son los siguientes:

- Factor 1: Historial delictivo.
- Factor 2: Educación formal/empleo.
- Factor 3: Financiero.
- Factor 4: Familiar/marital.
- Factor 5: Vivienda/alojamiento.
- Factor 6: Ocio/diversión.
- Factor 7: Red social.
- Factor 8: Abuso de alcohol/drogas.
- Factor 9: Salud mental.

➤ Factor 10: Actitudes/valores.

Una vez recogida la información, se vuelca el resultado de la evaluación, y se organiza y sistematiza la información de forma a ofrecer un resultado cualitativo y cuantitativo sobre el nivel de riesgo de reincidencia del recluso.

Predictive Data Justice busca la información requerida por el IFRIP en las fuentes de datos citadas proporcionando al funcionario de prisión una automatización en la búsqueda de dicha información.

5.6.2 Análisis descriptivo de las características de la muestra

En primer lugar, se realiza un estudio descriptivo de los aspectos demográficos y sociales, así como de algunos aspectos relacionados con el delito. Posteriormente, se aborda el estudio bivariado para el que se utiliza un estadístico paramétrico siempre que el tamaño muestral así lo permita. Se utilizará el estadístico **Chi-cuadrado** para analizar el cruce entre variables categóricas y la prueba “t” de Student para la comparación de dos medias, o la **ANOVA** cuando haya que comparar más de dos medias. Finalmente, se utilizarán técnicas de análisis multivariante para construir un modelo de propósito predictivo que medirá el efecto de una exposición en presencia de variables de confusión y de interacción.

5.6.3 Relación entre el análisis de la reincidencia delictiva y otros grupos de variables

En esta fase, se analiza la reincidencia delictiva en función de algunas de las variables contenidas en el fichero de datos SISPE. En el caso de un tamaño muestral pequeño (es decir, un número de observaciones inferior a 20) o de la presencia de frecuencias esperadas menores de 1, se aplicará la corrección por continuidad de Yates. Se consideran significativos aquellos estadísticos con una probabilidad de que la diferencia encontrada debida al azar es muy pequeña (inferior al 5%). Mediante la siguiente tabla se muestra un claro ejemplo de ello:

Tabla 5-2. Reincidencia general frente a tiempo de condena

Condena (años)	No		Sí		Total	
	Muestra	Porcentaje	Muestra	Porcentaje	Muestra	Porcentaje
< 1	38	5,9	0	0	38	4,7
1 - 5	230	41,4	104	40,6	334	41,2
5 - 10	203	36,6	99	38,7	302	37,2
10 - 15	57	10,3	34	13,3	91	11,2
15 - 20	27	4,9	19	7,4	46	5,7
Totales	555	100	256	100	811	100

5.6.4 Análisis del Inventario de Factores de Riesgo e Intervención en Prisiones (IFRIP)

En esta fase, se presenta el análisis estadístico del Inventario sobre Factores de riesgo e Intervención en Prisiones (IFRIP), con especial atención a sus propiedades psicométricas y a su validez predictiva. En cuanto a la validez predictiva del IFRIP, se utiliza **la curva ROC** para su estimación. La curva ROC es el resultado de la combinación de la tasa de verdaderos positivos (sensibilidad) frente a la tasa de falsos positivos (1 menos la especificidad) para cada punto de corte del instrumento, y asume un área bajo la curva (AUC) como predictor continuo.

El **área bajo la curva (AUC)** se define como la probabilidad de reincidir que tiene un recluso seleccionado aleatoriamente que puntúe alto en la medida de riesgo

específica (el IFRIP en nuestro caso) en comparación con otro recluso, también seleccionado aleatoriamente, que puntúe bajo.

Un área bajo la curva (AUC) de 0,50 representa una predicción realizada al azar, mientras que un AUC de 1 equivale a una predicción perfecta, es decir, el instrumento no se equivoca nunca a la hora de predecir la reincidencia delictiva. En el eje de ordenadas se representa la sensibilidad o fracción de verdaderos positivos. Es decir, la cantidad de reincidentes que la puntuación de la escala es capaz de identificar correctamente. En el eje de abscisas está la fracción de falsos positivos (1 menos la especificidad) que se calcula dentro del grupo de no reincidentes. En el supuesto de un AUC = 0,750, indica que si se extrae al azar un recluso del grupo de los que han puntuado alto en el IFRIP, este tiene una probabilidad del 75% de reincidir, valor muy superior al que obtendrá un recluso del grupo de los que han puntuado bajo en la prueba. Los segmentos diagonales de la curva ROC son producidos por los empates.

Tabla 5-3. Coordenadas de la curva ROC para la reincidencia delictiva

Positivo si es mayor o igual que (a)	Sensibilidad	1 - Especificidad
0,0268	1	0,995
0,0446	1	0,978
0,0625	0,996	0,959
0,0804	0,984	0,932
0,0982	0,984	0,91
0,1161	0,98	0,877
0,1339	0,973	0,838
0,1518	0,973	0,798
0,1696	0,93	0,699
0,1875	0,926	0,645
0,2054	0,918	0,596
0,2232	0,902	0,551

0,2411	0,863	0,505
0,2589	0,836	0,472
0,2768	0,813	0,429
0,2946	0,781	0,384
0,3125	0,754	0,341
0,3304	0,73	0,295
0,3482	0,703	0,267
0,3661	0,66	0,245
0,3839	0,617	0,225
0,4018	0,57	0,2
0,4196	0,512	0,18
0,4375	0,453	0,153
0,4554	0,402	0,121
0,4732	0,359	0,092
0,4911	0,328	0,077
0,5089	0,266	0,059
0,5268	0,227	0,05
0,5446	0,188	0,04
0,5625	0,141	0,032
0,5804	0,121	0,025
0,5982	0,078	0,018
0,6161	0,066	0,018
0,6339	0,055	0,013
0,6518	0,039	0,009

0,6696	0,031	0,005
0,6875	0,023	0,004
0,7054	0,016	0,002
0,7321	0,008	0,002
0,7589	0,004	0
0,7679	0	0

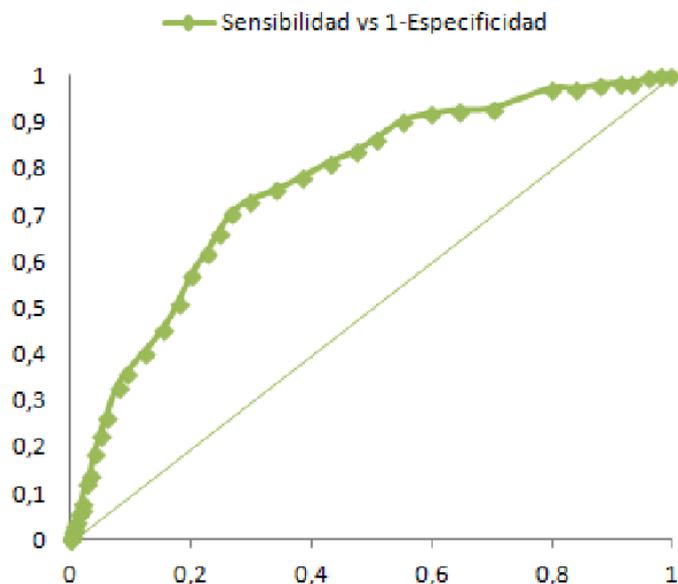


Figura 5-6. Curva ROC

En realidad, el valor del área bajo la curva (AUC) obtenido (igual a 0.77) corresponde a una puntuación del IFRIP que mejor permite clasificar a los reclusos reincidentes. Por tanto, se busca una puntuación que no comprometa la sensibilidad del instrumento y su especificidad y que, consecuentemente, detecte correctamente a los reclusos que reincidirán pero que al mismo tiempo no dé falsos positivos. **En este estudio se ha estimado que el mejor punto de corte para el IFRIP es de 0,31 ya que clasifica correctamente a un 75% de los reincidentes, ofreciendo un 34% de falsos positivos.**

5.6.5 Modelos explicativos de la reincidencia delictiva

Para esta última fase, se emplea la técnica de los **Modelos de Ecuaciones Estructurales (AMOS)** para evaluar la interrelación entre los factores del IFRIP y la reincidencia delictiva, combinada con la utilización de un **Árbol de Decisión** para establecer el riesgo de reincidencia delictiva (individualizado por recluso), cuyo conjunto de reglas se corresponde con los propios factores IFRIP.

Los **Modelos de Ecuaciones Estructurales (MES)** es una técnica estadística multivariante para probar y estimar relaciones causales a partir de datos estadísticos y asunciones cualitativas sobre la causalidad.

Esta técnica combina el análisis factorial con la regresión lineal para probar el grado de ajuste de unos datos observados a un modelo hipotético y expresado mediante un diagrama de senderos. Como resultado, los MES proporcionan los valores pertenecientes a cada relación, y más importante, un estadístico que expresa el grado en el que los datos se ajustan al modelo propuesto, confirmando su validez.

Entre los puntos fuertes de los MES se encuentra la habilidad de construir variables latentes: variables que no son medidas directamente, pero que son estimadas en el modelo a partir de varias variables que covarían entre sí. Esto permite al modelador capturar explícitamente la fiabilidad del modelo.

Todos los factores del IFRIP correlacionan significativamente con la reincidencia delictiva, a excepción del factor 6 (ocio/diversión). Los dos factores más asociados a este tipo de reincidencia son el factor 1 (historial delictivo) y el factor 8 (abuso de alcohol/drogas).

Tabla 5-4. Correlaciones entre cada uno de los factores del IFRIP y la reincidencia delictiva

Factores del IFRIP	Reincidencia delictiva
Factor 1 (Historial delictivo)	0,559, con probabilidad inferior a 0,001
Factor 2 (educación/empleo)	0,171, con probabilidad inferior a 0,005
Factor 3 (Financiero)	0,180, con probabilidad inferior a 0,005

Factor 4 (Familiar/marital)	0,158, con probabilidad inferior a 0,005
Factor 5 (Vivienda/alojamiento)	0,126, con probabilidad inferior a 0,005
Factor 6 (Ocio/diversión)	0,02
Factor 7 (Red social)	0,282
Factor 8 (Abuso de alcohol/drogas)	0,353, con probabilidad inferior a 0,005
Factor 9 (Salud mental)	0,149, con probabilidad inferior a 0,005
Factor 10 (Actitudes/valores)	0,10, con probabilidad inferior a 0,005

El método de estimación a utilizar es el **método Asintóticamente Libre de Distribución (ADF)** que permite al modelo ser eficiente para cualquier distribución de las variables observadas. Este método de estimación permite obtener estimaciones adecuadas de los modelos sin el requisito de la distribución normal de las variables utilizadas para determinar la bondad de ajuste del modelo propuesto, suelen ser adecuados los valores superiores a 0,90 tanto en el denominado “Índice de Bondad de Ajuste” (Goodness of Fit Index, o GFI) como en el “Índice Ajustado de Bondad de Ajuste” (Adjusted Goodness of Fit Index, o AGFI). En el caso del “Error cuadrático medio de aproximación” (Root Mean Square Error of Approximation, o RMSEA) los valores aproximados 0,05 suelen indicar un buen ajuste del modelo.

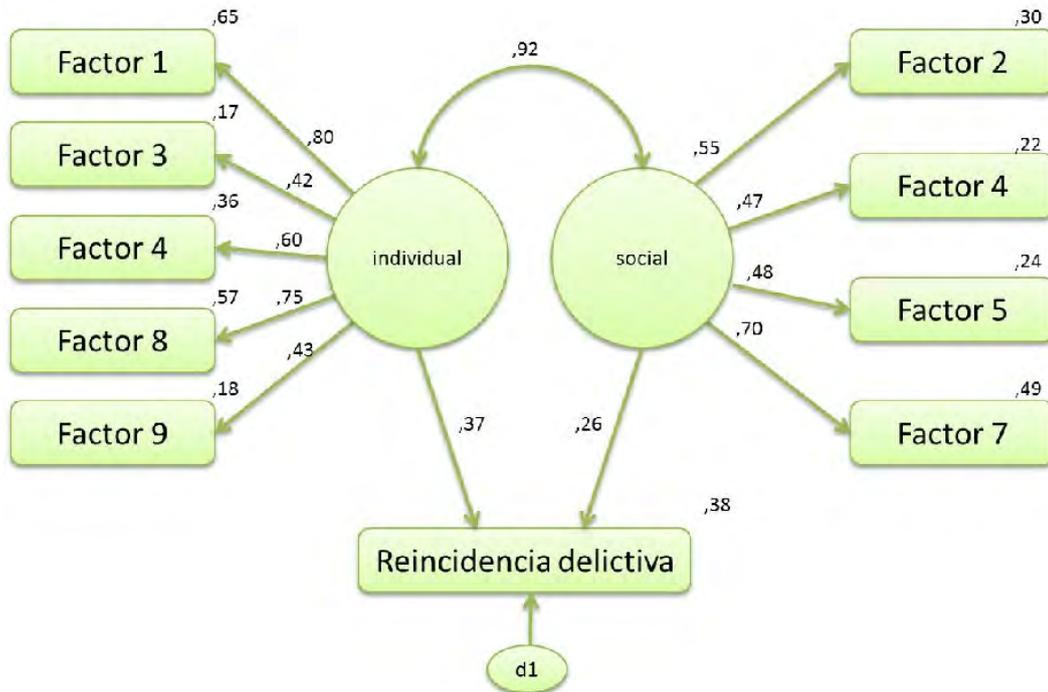


Figura 5-7. Modelo explicativo de la reincidencia delictiva

Las varianzas explicadas de cada factor del IFRIP se representan en el margen superior derecho del factor. Los coeficientes estructurales se representan al lado de cada una de las flechas presentes en el modelo.

Específicamente en este modelo, se observa que **los factores 1, 3, 6, 8, y 9**, quedan agrupados satisfactoriamente en el constructo denominado “individual” al agrupar factores de riesgo que tienen que ver con factores personales o individuales en la reincidencia delictiva. Asimismo, **los factores 2, 4, 5, y 7** quedan agrupados en el constructo “social” dado que todos estos factores hacen alusión a factores de naturaleza social e interpersonal. Todos los factores individuales y sociales presentan una varianza explicada satisfactoria, desde un 0,18 (del factor 9) hasta un 0,57 (del factor 8).

Dentro de los factores individuales, el factor con mayor contribución relativa es el factor 8 (con un 0,75), mientras que el factor 7 es el que más contribución relativa presenta en los factores sociales (con un 0,70). Ambos constructos, individual y social, presentan una correlación significativa y elevada entre sí (de 0,92).

En consecuencia, la reincidencia general delictiva queda explicada satisfactoriamente por el modelo planteado (varianza de 0,38) y los coeficientes estructurales están significativamente asociados a la reincidencia, tanto el individual (con un coeficiente de 0,37; con probabilidad inferior a 0,05) como el

social (con un coeficiente de 0,26; con probabilidad inferior a 0,05); por lo que los factores individuales del IFRIP tienen una mayor contribución relativa en la predicción de la reincidencia delictiva. La Figura 5-6 presenta el modelo de ecuaciones estructurales entre la violencia delictiva y los respectivos factores del IFRIP. Los índices obtenidos (GFI = 0,96, AGFI = 0,93, y RMSEA = 0,06) vuelven a indicar que este modelo también presenta una adecuada bondad de ajuste entre las relaciones de los factores del IFRIP planteadas y los datos empíricos contrastados.

5.7 Motor de recomendaciones

El Know How de **Predictive Data Justice** es la mezcla de la implementación del modelo predictivo de reincidencia con su motor de recomendaciones. Por definición, consiste en un conjunto de metodologías y algoritmos que permiten afrontar científicamente los procesos de personalización, y en base a ello mejorar los procesos de toma de decisiones.

Más concretamente, se trata de un caso especial de **NBA (Next Best Action)** para apoyar a la toma de decisiones, que como paradigma centralizado en el recluso, tratará de establecer un ranking con las diferentes acciones propuestas que aplicadas sobre el mismo, consigan reducir su riesgo de reincidencia delictiva.

Dichas acciones forman en su conjunto un completo catálogo ofertado por todas las organizaciones, fundaciones y demás entidades colaboradoras de la SGIIPP, que son introducidas al sistema a través de un denominado “**gestor de acciones**” por uno o varios usuarios con permisos específicos para ello, por medio de un proceso de carga en modo batch.

Cada acción que se introduzca en el sistema, está caracterizada por un código identificativo y descriptivo de la misma, junto con los datos relativos a la entidad colaboradora de la SGIIPP que la desarrolla, personas de contacto, su coste total y el grado de efectividad previa en su aplicación a reclusos.

Además, el motor de recomendaciones se nutre de todos los resultados que se desprenden del modelo analítico-descriptivo de evaluación para el riesgo de reincidencia delictiva, comentado anteriormente a lo largo del presente capítulo. Dichos resultados son tratados por un modelo que analiza la información disponible sobre cada recluso, y obtiene una lista de ordenada y personalizada de posibles acciones para cada tipo de recluso, con el fin de que cualquier usuario de la solución con los permisos requeridos para ello, pueda asignar al recluso solo

aquellas acciones que sean las más adecuadas para lograr su completa reeducación y reinserción social.

En base a nuestra experiencia previa, el modelo implantado en el motor de recomendaciones se basa en una **regresión logística**, puesto que se trata de un modelo muy conocido y cuya utilización está bastante extendida en problemas relacionados con la evaluación del riesgo. Además, tenemos constancia de que sigue funcionando muy bien incluso si no se cumplen todos los supuestos, y se puede simular con cierta facilidad.

5.8 Visualización

Continuando con lo explicado en el capítulo 5.5, la maquetación visual de la aplicación tendría dos perfiles, el perfil de operario (funcionario de prisiones o educador) que realiza el seguimiento del recluso y un perfil de mayor nivel que especifica las reglas del motor de recomendaciones.

Inicialmente la aplicación que se aloja en un servidor de aplicaciones WEB dentro de las instalaciones del cliente, muestra una **pantalla de inicio / login**. Una vez que el usuario se haya registrado con su nombre y contraseña, puede acceder al buscador o directamente ir a la pantalla de modificaciones recientes.

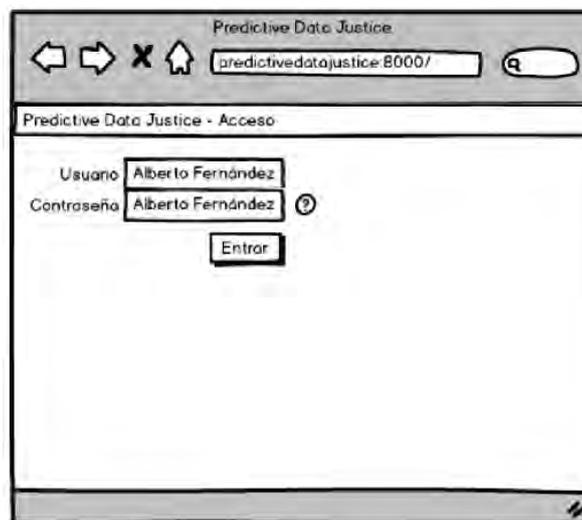


Figura 5-8. Login de usuario

El **buscador de texto natural** le permite al usuario buscar a cualquier recluso del establecimiento penitenciario y poder acceder al perfil del mismo en función de los permisos asociados.



Figura 5-9. Buscador de recluso

Una vez seleccionado el recluso, el funcionario de prisión puede ver su perfil que se compone de los siguientes apartados:

- **Datos personales del recluso.** Los datos personales son extraídos de fuentes de datos estructurados y se actualizan una vez al día (no existe federación). Son datos con una gran latencia en la caducidad y suelen no ser fácilmente alterables. Estos datos están especialmente protegidos por la Ley Orgánica de Protección de Datos.
- **Datos penitenciarios.** Todo el seguimiento del recluso que circula en la prisión se refleja en informes que generan los funcionarios de prisión, educadores y psicólogos. El sistema permite anexarlos o incluso obtenerlos del SISPE con el objeto de clasificarlos mediante esta pestaña, vista o apartado.
- **Datos históricos judiciales.** Del mismo modo, el sistema recupera de forma frecuente y en batch los documentos PDF generados por el proceso judicial. Este visor de datos judiciales permite al funcionario conocer el detalle de la información procesal del preso.

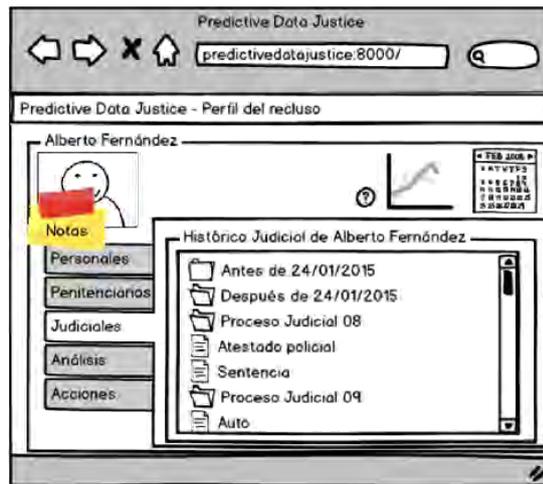


Figura 5-10. Pestaña del histórico judicial

- **Análisis de reincidencia delictiva.** Gracias al modelo predictivo calculado en función de las técnicas descritas en capítulos anteriores, el sistema es capaz de obtener un análisis parametrizado del riesgo de reincidencia del recluso seleccionado con el fin de generar un mapa de acciones de reinserción.



Figura 5-11. Análisis de riesgo de reincidencia

- **Seguimiento de las acciones.** Esta pestaña permite visualizar y realizar un seguimiento así como obtener feedback del proceso de ejecución de las acciones propuestas en el anterior punto.

Predictive Data Justice es un producto fácil de personalizar, todos los "mock ups" presentados en este capítulo son un ejemplo de las funcionalidades ofrecidas por el producto. En el **Anexos**

Centros Penitenciarios

La **distribución de los** diferentes centros penitenciarios en el territorio nacional es la que se muestra en la siguiente figura.



Figura 13-1. Centros Penitenciarios

La distribución por centros territoriales para esta primera fase de implantación incluye las siguientes regiones:

- CT8. Castilla-La Mancha y Extremadura.
- CT5. Comunidad Valenciana, Murcia e Islas Baleares.
- CT10. Sevilla, Cadiz y Huelva.

- CT8. Asturias, Cantabria y País Vasco.
- CT4. Comunidad de Madrid, únicamente los centros penitenciarios:
 - Madrid I. Mujeres. Alcalá de Henares.
 - Madrid II. Alcalá de Henares.
 - Madrid III. Valdemoro.
 - Madrid IV. Navalcarnero.

5.9 Criterios de evaluación de riesgos

En las siguientes tablas se encuentran las definiciones que se han utilizado para realizar una evaluación de los riesgos:

Entrada	Definición
Identificador	Identificador unívoco del riesgo. Este identificador estará compuesto por un contador incremental según la estructura PDJ-RIE.[XXXX].
Descripción	Breve descripción del riesgo, indicando sus posibles causas así como una descripción de los perjuicios que puede causar
Impacto	Se debe indicar el impacto del riesgo en el proyecto
Probabilidad	Se debe indicar la probabilidad de ocurrencia del riesgo en el proyecto.
Criticidad	Este campo es establecido en función del impacto y la probabilidad.
Estrategia	Refiere la principal estrategia, o combinación de estrategias para la gestión de riesgos. Los posibles valores serán: eliminar, mitigar, aceptar, transferir.
Acciones	Se indican las diversas acciones que se tomarán para su gestión.

Valoración del impacto del riesgo

Valor	Definición
Muy alto	Su ocurrencia afectará a objetivos críticos del proyecto no se cumplirán (coste, calendario, calidad o satisfacción del cliente), pudiendo incluso causar la no viabilidad del proyecto
Alto	Su ocurrencia implicará serios y significativos perjuicios en el proyecto (retrasos, aumentos de costes, modificación en el alcance) y afectará a objetivos críticos del proyecto
Medio	Su ocurrencia implicará perjuicios importantes en el proyecto
Bajo	Su ocurrencia implicará sobreesfuerzos asumibles en el proyecto

Probabilidad de ocurrencia del riesgo

Valor	Definición
Muy alta	La probabilidad es cercana a la seguridad de ocurrencia (>0.9)
Alta	La probabilidad de ocurrencia es muy elevada (0.6-0.9)
Media	La probabilidad de ocurrencia es moderada (0.3-0.6)
Baja	Difícilmente o muy difícilmente se materializará el riesgo (<0.3)

Estrategia

Valor	Definición
Evitar	Implica tomar todas las medidas necesarias para la eliminación total de la amenaza
Transferir	Implica trasladar a un tercero todo o parte del impacto negativo de una amenaza, junto con la propiedad de respuesta. Esto confiere a una tercera persona la propiedad de la responsabilidad de su gestión; pero no lo elimina
Mitigar	Implica tomar las medidas necesarias para reducir a un umbral (nivel) aceptable la probabilidad y/o el impacto del evento. Se consigue aplicando técnicas de prevención, que ayuden a reducir la probabilidad de ocurrencia del riesgo. En aquellos casos, que no pueda reducirse la probabilidad de ocurrencia, deberá de dirigir los esfuerzos en reducir la severidad del mismo, es decir, añadir contramedidas que reduzcan el impacto del mismo
Aceptar	Se adopta la técnica de aceptar los riesgos, de dos maneras: activas o pasiva. La activa corresponde a disponer de un plan de contingencias para hacer frente a la amenaza, y la pasiva en dejar que el equipo de trabajo gestione el riesgo conforme se presenta

Mock-ups de la aplicación visual, se detallan todas las pantallas que el usuario final funcionario de prisiones utilizará después de la implantación de Predictive Data Justice.

6 Plan de proyecto

6.1 Metodología

		Áreas de Conocimiento								
		1. Gestión de la Integración del Proyecto	2. Gestión del Alcance del Proyecto	3. Gestión del Tiempo del Proyecto	4. Gestión de Costes del Proyecto	5. Gestión de la Calidad del Proyecto	6. Gestión de los RRHH del Proyecto	7. Gestión de las Comunicaciones del Proyecto	8. Gestión de Riesgos del Proyecto	9. Gestión de Adquisiciones del Proyecto
Grupos de Procesos de la Dirección de Proyectos	Procesos de Iniciación	1.1. Desarrollar el acta de Constitución del Proyecto						7.1. Identificar a los Usuarios		
	Procesos de Planificación	1.2. Desarrollar el Plan para la Dirección del Proyecto	2.1. Recopilar requisitos 2.2. Definir el alcance 2.3. Crear el WBS	3.1. Definir las actividades 3.2. Secuenciar las actividades 3.3. Estimar los Recursos de las actividades 3.4. Estimar la duración de las actividades 3.5. Desarrollar el cronograma	4.1. Estimar los costes 4.2. Determinar el Presupuesto	5.1. Planificar la Calidad	6.1. Desarrollar el Plan de Recursos Humanos	7.2. Planificar las Comunicaciones	8.1. Planificar la Gestión de Riesgos 8.2. Identificar los Riesgos 8.3. Realizar el análisis cualitativo de Riesgos 8.4. Realizar el Análisis Cuantitativo de Riesgos 8.5. Planificar la Respuesta a los Riesgos	9.1. Planificar las Adquisiciones
	Procesos de Ejecución	1.3. Dirigir y Gestionar la Ejecución del Proyecto				5.2. Realizar el Aseguramiento de la Calidad	6.2. Adquirir el Equipo de Proyecto 6.3. Desarrollar el Equipo de Proyecto 6.4. Gestionar el Equipo de Proyecto	7.3. Distribuir la información 7.4. Gestionar las expectativas de los usuarios		9.2. Efectuar las Adquisiciones
	Procesos de Seguimiento y Control	1.4. Monitorear y controlar el trabajo del proyecto 1.5. Realizar el Control Integrado de Cambios	2.4. Verificar el Alcance 2.5. Controlar el Alcance	3.6. Controlar el Cronograma	4.3. Controlar los Costes	5.3. Realizar el Control de Calidad		7.5. Informar del Avance	8.6. Monitorear y Controlar los Riesgos	9.3. Administrar las Adquisiciones
	Procesos de Cierre	1.6. Cerrar el Proyecto o Fase								9.4. Cerrar las Adquisiciones

Para implantar nuestro proyecto vamos a adaptar el estándar internacional de **metodología para Gestión y Dirección de Proyectos PMP**. Dicha metodología se divide en los diferentes procesos y áreas de conocimiento como se reflejan en la figura de la página anterior. En **Predictive Data** trabajamos dicha matriz a la hora de configurar la metodología y la planificación como se detalla a continuación en las siguientes páginas.

El enfoque metodológico se ha establecido en las siguientes fases, teniendo en cuenta la experiencia en otras implantaciones dentro del área Business Analytics.

- Inicio de proyecto con un Kick of meeting.
- Análisis y planificación.
- Diseño y construcción. Desarrollo y verificación.
- Validación, despliegue y aceptación del sistema.
- Soporte y mantenimiento.



Figura 6-1. Inicio del proyecto



Figura 6-2. Análisis y planificación



Figura 6-3. Diseño y construcción



Figura 6-4. Validación y despliegue



Figura 6-5. Soporte y mantenimiento

6.2 Implantación

6.2.1 Planificación de tareas

Para la implantación del proyecto **Predictive Data Justice** se han definido las siguientes fases basados en la metodología detallada en el punto anterior con la siguiente planificación a alto nivel como línea base del proyecto

A medida que avance el proyecto se irán definiendo cada una de las tareas en detalle, como corresponde al proceso de Seguimiento y Control dentro de la Gestión Integral de Proyectos de la metodología PMP.

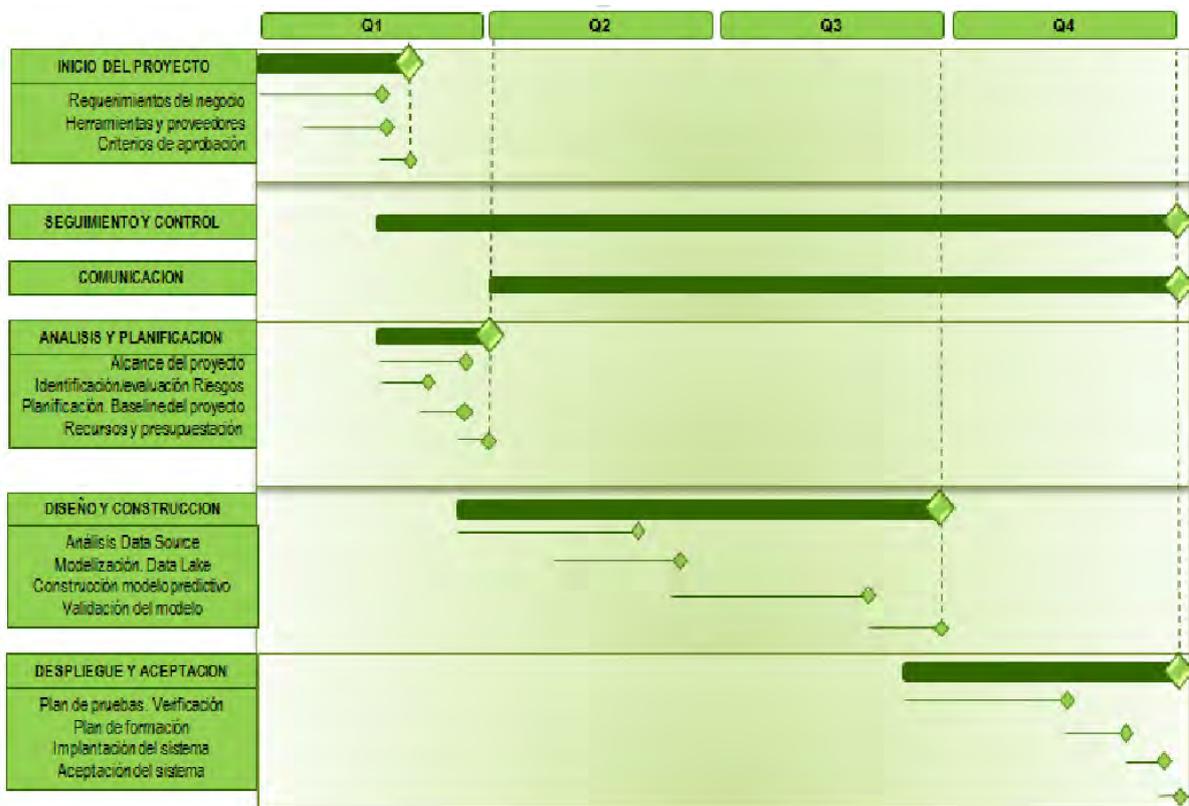


Figura 6-6. Cronograma

6.2.2 Fases de implantación de Predictive Data Justice

Inicialmente se ha estimado un periodo de tres años para la implantación del producto. **Durante el transcurso del cuarto año, es cuando se deben medir la rentabilidad y eficiencia de la reincidencia delictiva de Predictive Data Justice** y su verdadero potencial. Su implantación se va a realizar en parte del territorio nacional de forma estratégica, asegurando que los objetivos se cumplan en algunos centros y se **comparen con otros territorios**.

Para la selección de los centros, se ha tenido en cuenta dos factores importantes:

➤ **Distribución de las Coordinaciones Territoriales (CT) para trabajos penitenciarios.**

Para que el sistema sea eficiente, es importante que compartan dicha distribución para la elaboración de los programas y/o trabajos de reinserción. Dicha distribución está representada en el siguiente mapa.

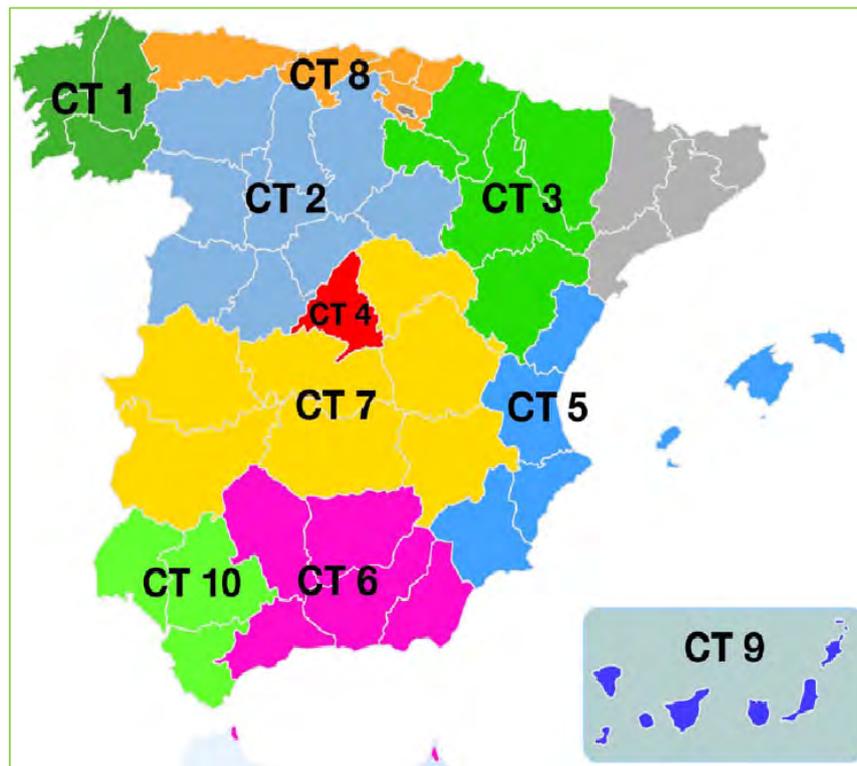


Figura 6-7. Distribución territorial Centros Penitenciarios

➤ **Estadística población reclusa**

Se han recogido datos del Anuario Estadístico del Ministerio del Interior correspondiente al año 2014 para valorar aquellos centros estratégicos donde Predictive Data Justice aportaría más valor al objetivo que se persigue.

Los compromisos de implantación incluyen los siguientes centro penitenciarios agrupados por Coordinaciones Territoriales (CT). Ver **13.1 Centros Penitenciarios**

Tabla 6-1. Plan de implantación

Áreas y/o centros de implantación	2015		2016		2017	
	S1	S2	S1	S2	S1	S2
CT7. Castilla-La Mancha y Extremadura						
CT5. Comunidad Valenciana, Murcia e Islas Baleares						
CT10. Sevilla, Cádiz y Huelva						
CT8. Asturias, Cantabria y País Vasco						
CT4. Comunidad de Madrid (*)						

(*) Los centros de Madrid I. Mujeres. Alcalá de Henares; Madrid II. Alcalá de Henares; Madrid III. Valdemoro y Madrid IV. Navalcarnero

En futuras fases, al ser una solución escalable se irán implantando en todo el territorio nacional, realizando el análisis y preprestación oportunos.

6.2.3 Riesgos

A la hora de implantar un proyecto, hay que evaluar ciertos riesgos que pueden desviar los plazos comprometidos y los acuerdos de servicio establecidos. Para ello se deben tomar ciertas estrategias y/o planes de contingencia que mitiguen o eviten dichos riesgos. En los que casos en que se acepten hay que tomar medidas para que no se produzcan dichas desviaciones.

Los riesgos identificados al inicio del proyecto, con sus medidas y planes de respuesta que además, han sido tenidos en cuenta para cuantificar los costes son los siguientes. (Ver **13.2 Criterios de evaluación de riesgos**).

Tabla 6-2. Riesgos iniciales identificados en el proyecto

Ident	Descripción	Impacto	Probabilidad	Criticidad	Estrategia	Acciones - Planes
PDJ-RIE.001	Cambios organizativos en las instituciones. Rotación de personal clave	Bajo	Alta	Medio	Aceptar	Preparar planes específicos para velar por la estabilización presupuestaria y cumplimiento de los objetivos establecidos inicialmente.
PDJ-RIE.002	Continuas solicitudes de cambio por parte de las instituciones o agentes externos (cambios de prioridad, inclusión de nuevos trabajos, etc.)	Medio	Alto	Alta	Mitigar	Estimarlas correctamente para adaptarlas al Plan de Proyecto y solicitar su priorización por parte de la Institución
PDJ-RIE.003	Deficiencias de coordinación entre los diferentes sistemas de información. Complejidad integración de las fuentes	Alto	Bajo	Medio	Mitigar	Gestión de los recursos para elaborar un plan detallado de las diversas fuentes de información, volumetría e interoperabilidad entre sistemas.
PDJ-RIE.004	Falta de definición en el alcance y requisitos de algunas de las peticiones de los usuarios de la AP	Bajo	Baja	Bajo	Evitar	Reuniones con el usuario para delimitar el alcance y requisitos de esas peticiones
PDJ-RIE.005	Incertidumbre en las partidas presupuestarias al inicio del ejercicio dentro del programa para cubrir otras necesidades de la Administración	Medio	Baja	Medio	Aceptar	Seguimiento y control del presupuesto y revisión de los acuerdos de servicio acordados con la Administración
PDJ-RIE.006	Estimación errónea de la planificación para la implantación y el despliegue	Alta	Baja	Alta	Evitar	Seguimiento de la línea base del proyecto y elaborar las medidas correctoras necesarias para el cumplimiento de los compromisos establecidos con la Administración

A lo largo del proyecto, tal y como dicta la metodología PMP, pueden aparecer otros riesgos no identificados en su fase inicial. Éstos deben ser registrados y controlados por la persona gestora de ello, para evitar desviaciones que no materialicen los objetivos acordados.

7 Prueba de concepto

Después de haber definido el problema de negocio y la solución, un elemento clave de cohesión para el proyecto es la **prueba de concepto** que nos ha servido para afianzar la solución y definir los requisitos de la misma.

En el capítulo 3.3, **Contactos y entrevistas**, se detallan los pasos que hemos seguido para corroborar la viabilidad de nuestro proyecto desde un punto de vista de cliente y usuario final.

Por otro lado y después de haber estudiado en detalle el modelo y sus fuentes de datos para que la solución sea viable hemos decidido realizar algunas pruebas de

Minería de Texto utilizando el paquete TM del lenguaje R. En este capítulo se detallan las pruebas realizadas utilizando datos reales del Ministerio de Justicia obtenidos del CENDOJ.

Todos los ficheros de **lenguaje R** que han sido implementados se encuentran anexados como apéndice de la memoria, **Código fuente en R**.

7.1 Entorno de las pruebas

Las pruebas han sido realizadas con el entorno de desarrollo (IDE) **Open Source RStudio**. La versión de R utilizada ha sido la 3.1.2.

Los datos de entrada utilizados han sido sentencias y autos reales obtenidas del CENDOJ, descargadas en formato PDF. El total de documentos utilizados para la realización del ejercicio han sido 250.

El objetivo de las pruebas es únicamente la utilización de R sobre texto para comprobar el comportamiento del mismo.

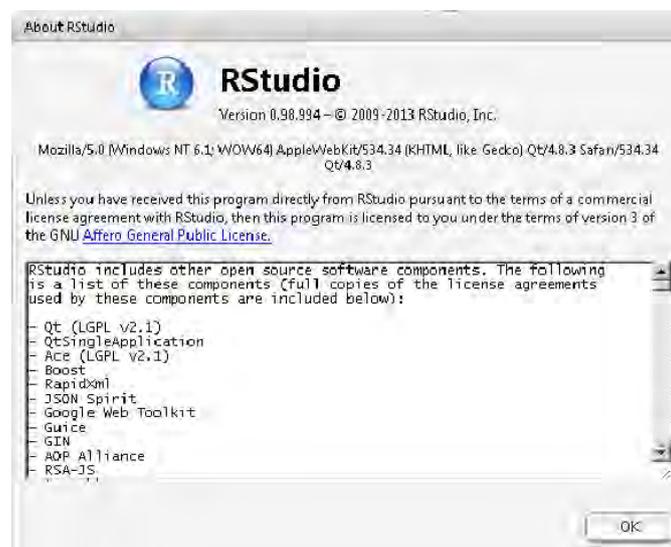


Figura 7-1. Ejemplo de RStudio

Todas las pruebas se han ejecutado con la misma base de minería de texto soportada por el paquete TM de R. El proceso es el siguiente.

- Descargar la información a una ruta local.
- Leer la ruta local y buscar todos los ficheros PDF.

- De forma automática generar ficheros de texto a partir de los ficheros en PDF.
- Cargar un Corpus (Corpora) de TM para todos los ficheros.
- Limpiar el contenido del Corpus eliminando los números y signos de puntuación, espacios en blanco, palabras en español no relevantes (determinantes, pronombres, etc.), utilizar las raíces de las palabras como índices coincidentes de una misma entrada en el cálculo de frecuencias.

7.2 Nube de Tags

Una vez obtenido el Corpus de las sentencias y autos hemos transformado el Corpus en una matriz con frecuencias de aparición por concepto y lo hemos pintado por tamaño.

El siguiente paso es identificar las palabras que se repitan mucho y generar una lista de estas palabras que no interesen y poder limpiar de contenido repetitivo el Corpus de texto.

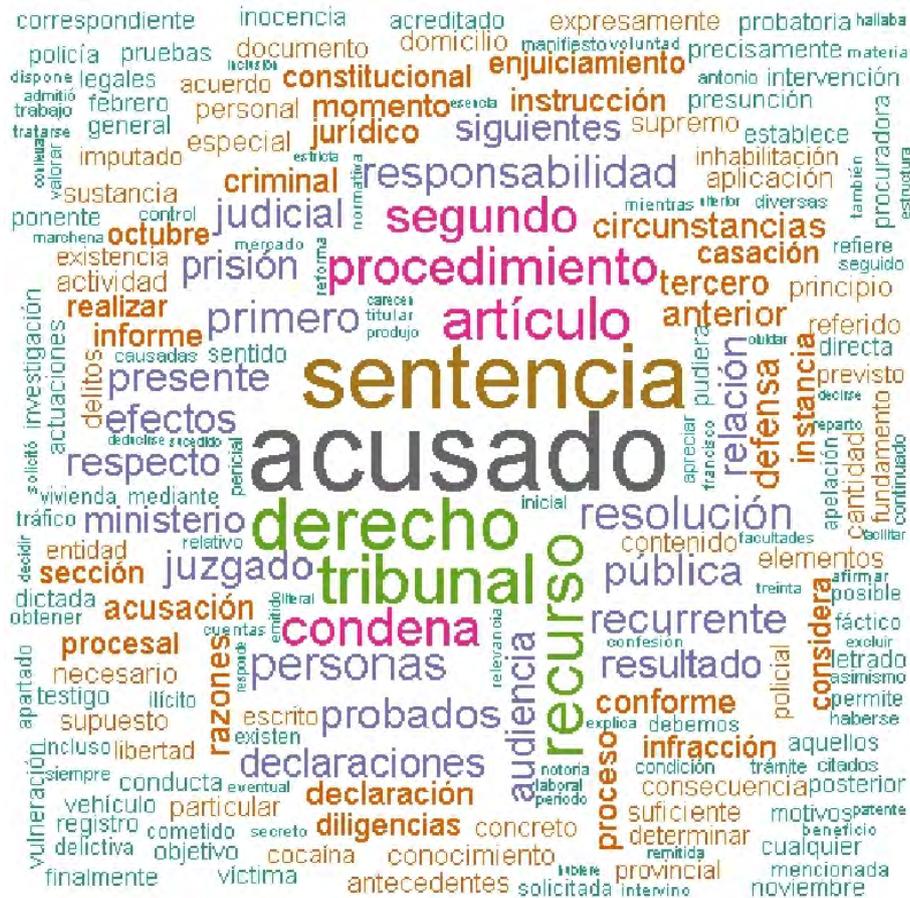


Figura 7-2. Resultado de la ejecución de la nube de tags

7.3 Dendograma y clustering sobre texto

Del mismo modo que el anterior caso, se ha transformado el Corpus en una Matriz y a su vez observando cómo de disperso se encontraban los elementos indexados en el Corpora, se decide decrementar esta dispersión con el objetivo de afinar en la clusterización.

```
> mydata.dtm
<<TermDocumentMatrix (terms: 31617, documents: 243)>>
Non-/sparse entries: 240575/7442356
Sparsity           : 97%
Maximal term length: 33
Weighting          : term frequency (tf)
> mydata.dtm2
<<TermDocumentMatrix (terms: 74, documents: 243)>>
Non-/sparse entries: 15399/2583
Sparsity           : 14%
Maximal term length: 14
Weighting          : term frequency (tf)
> |
```

Figura 7-3. Modificación del atributo Sparsity en la Matriz

Para reducir el **Sparsity** se utiliza la función de TM "removeSparseTerms()" obteniendo un 14% de dispersión en la muestra. Posteriormente se aplica un análisis de clúster jerárquico basado en el método de Ward (Joe H. Ward, Jr), de varianza mínima utilizando el error de la suma de los cuadrados. Se puede observar el resultado visual en la siguiente página.

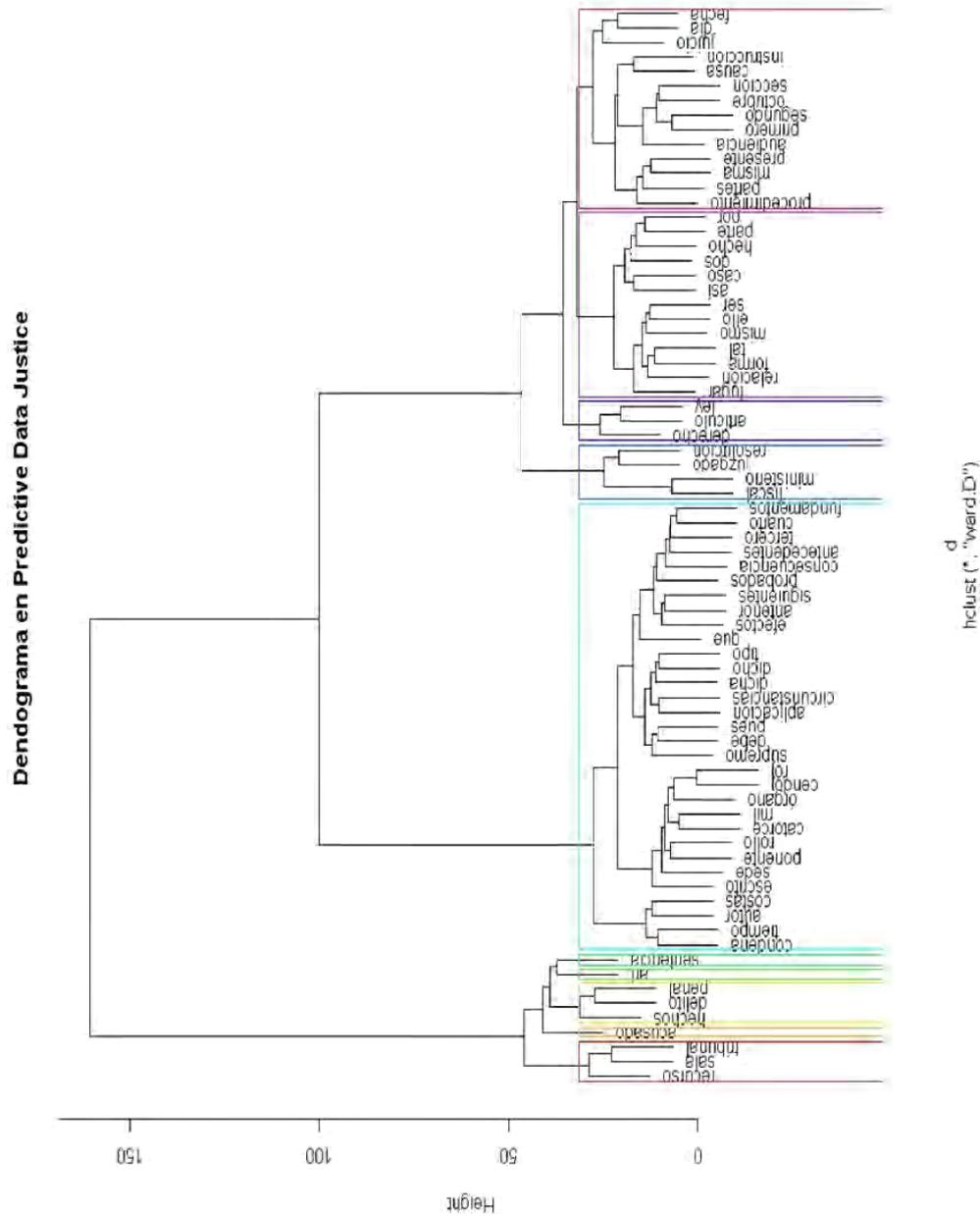


Figura 7-4. Dendrograma con 10 clústeres ejecutado sobre sentencias

7.4 Relaciones de reincidencia

Igualmente y por último, con los conocimientos obtenidos en las anteriores pruebas, se ha decidido jugar con los datos comprobando que existe cierta relación con las estadísticas mostradas en el capítulo **2.2 Comparativa de España frente a Europa**. Se ha utilizado la función "findAssocs" con el objetivo de encontrar las asociaciones entre conceptos con el siguiente resultado.

	antecedentes
penales	0.83
telefónica	0.62
intervención	0.61
segundo	0.61
costas	0.60
drogas	0.60
oficio	0.60
comunicaciones	0.59
siguientes	0.59

Figura 7-5. Dendrograma con 10 clústeres ejecutado sobre sentencias

En la siguiente lista y omitiendo las palabras repetitivas y que no intervienen en el proceso de ingesta porque son descartadas mediante removeStopWords (función implementada por nosotros), se puede observar la relación del concepto "**antecedentes**" con el concepto "**drogas**", claramente, esto avala tanto el estudio estadístico presentado así como el modelo predictivo basado en los factores identificados en el estudio.

8 Business case

Con **Predictive Data Justice** estamos convencidos de que se direccionarán de forma inteligente todas aquellas **partidas presupuestarias destinadas a la reinserción de los presos**, coordinadas con medidas poseedoras de una alta tasa de eficiencia en la reducción del comportamiento delictivo de los reclusos.

La mejora del tratamiento de la información de las bases de datos jurídicas-penitenciarias, reducirá los costes de reinserción con eficiencia, mejorará la reinversión de los fondos destinados a tal fin, y paliará la reincidencia delictiva.

El análisis de pre-viabilidad económico-financiera de nuestro proyecto de inversión, constituye una etapa imprescindible a la hora de tomar la decisión de comprometer una serie de recursos monetarios. El explícito y **fuerte valor social** de este proyecto, hace que nos centremos en el análisis **desde el punto de vista de la Administración**, analizando su problemática como un proyecto de servicio público, contemplando nuestra colaboración público-privada como un **estudio de pre-viabilidad**.

Para ello tomamos como punto de partida los **Presupuestos Generales del Estado (PGE) de 2015** destinado a **Instituciones Penitenciaria**. La partida presupuestaria es de **1.124.663.000,95 €**.

El actual porcentaje medio de reincidencia delictiva ronda el 55%, aunque depende de la tipología del delito. Diversos estudios e informes, como el presentado por el **Colegio Oficial de Psicólogos de Madrid**, dictaminan que siguiendo fehacientemente el criterio de los informes psicológicos, en algunos tipos de delitos se podría reducir hasta en un 44%. Del mismo modo en Chile una política acertada de reinserción social podría reducir hasta un 32%.

Teniendo en cuenta que nuestro modelo analítico **ha clasificado correctamente un 75% de los reincidentes, ofreciendo un 34% de falsos positivos, como se ha descrito en el punto 5.6.4 Análisis del Inventario de Factores de Riesgo e Intervención en Prisiones (IFRIP)**, **las reducciones medias previstas de la tasa de reincidencia en los próximos 4 años-** teniendo en cuenta que se implanta en parte del territorio nacional a partir de 2016- son **3%, 7,5% y 16%** respectivamente.

Tabla 8-1. Reducción tasa de reincidencia

	Tasa de reincidencia	
	Reducción	Tasa final
2015 - Año 1		55%
2016 - Año 2	3%	52%
2017 - Año 3	7,50%	44,50%
2018 - Año 4	16%	28,50%

8.1 Inversión y costes

La inversión que debe realizar la Administración Pública en la implantación de **Predictive Data Justice** durante los 4 próximos años en los centros o coordinaciones territoriales que se han detallado en el punto **6.2.2 Fases de implantación de Predictive Data Justice** se reflejan en el siguiente cuadro.

Tabla 8-2. Costes de implantación Predictive Data Justice

Costes de Implantación	2015	2016	2017	2018	Totales
Equipamiento Hw	116.100	35.100	31.300		182.500
Redes y Comunicaciones	58.400	40.000	22.000		120.400
Consultoría	377.320	350.840	256.760		984.920
Soporte y Mantenimiento		62.920	55.600	193.240	311.760
Formación	7.100	7.900	16.150		31.150
Otros (*)	4.150	4.150	4.650	2.150	15.100
Total por periodo	563.070	500.910	386.460	195.390	1.645.830

(*) Alojamiento y Transporte

Dicha inversión asciende a **1.645.830 €** en nuestro producto, por lo que la inversión inicial es relativamente baja con respecto al retorno que se puede llegar a obtener, no sólo **económico** sino a nivel **social**, posee un bajo coste de mantenimiento, y es la mejor solución tecnológica actualmente existente.

Por otra parte, en el análisis para la Administración se ha tenido en cuenta dentro de las partidas presupuestarias, los **costes para programas y trabajos de reinserción**. Estos costes son variables y varían en función de la reducción de reincidencia delictiva que proporciona la implantación de **Predictive Data Justice**.

Tabla 8-3. Costes de la Administración en programas de reinserción

Costes Acciones Reinserción	2015	2016	2017	2018	Totales
Conciertos asistencia sanitaria	9.697.947	9.382.763	8.970.601	8.146.275	36.197.585
Trabajos penitenciarios. Formacion empleo	6.726.951	6.508.325	6.222.430	5.650.639	25.108.345
UNED	192.885	186.616	178.419	162.023	719.943
Total por periodo	16.617.783	16.077.705	15.371.449	13.958.937	62.025.873

8.2 Rentabilidad económica

Los resultados de la rentabilidad financiera son **enormemente positivos**, debido a la dimensión de las magnitudes que empleamos, explicando la adopción de este análisis de pre-viabilidad, analizando el servicio desde el punto de vista de la administración, que nos invita reflexionar sobre el enorme recorrido, beneficio e influencia que tiene la adopción de cualquier medida relacionada con este tipo de actividad, con gran **repercusión social**.

Tabla 8-4. Indicadores de rentabilidad

Indicador	Valor
ROI	54,28%
Pay-back	24 meses
VAN *	23.152.628 € * k=10% de tasa de interés
TIR	52,38%

Tabla 8-5. Resumen flujo neto

	Ingresos	Costes e inversiones	Flujo Neto
2015 - Año 1		17.393.813	-17.393.813
2016 - Año 2	20.103.351	16.694.495	3.408.857
2017 - Año 3	26.288.998	15.902.909	10.386.089
2018 - Año 4	52.577.995	14.160.097	38.417.898

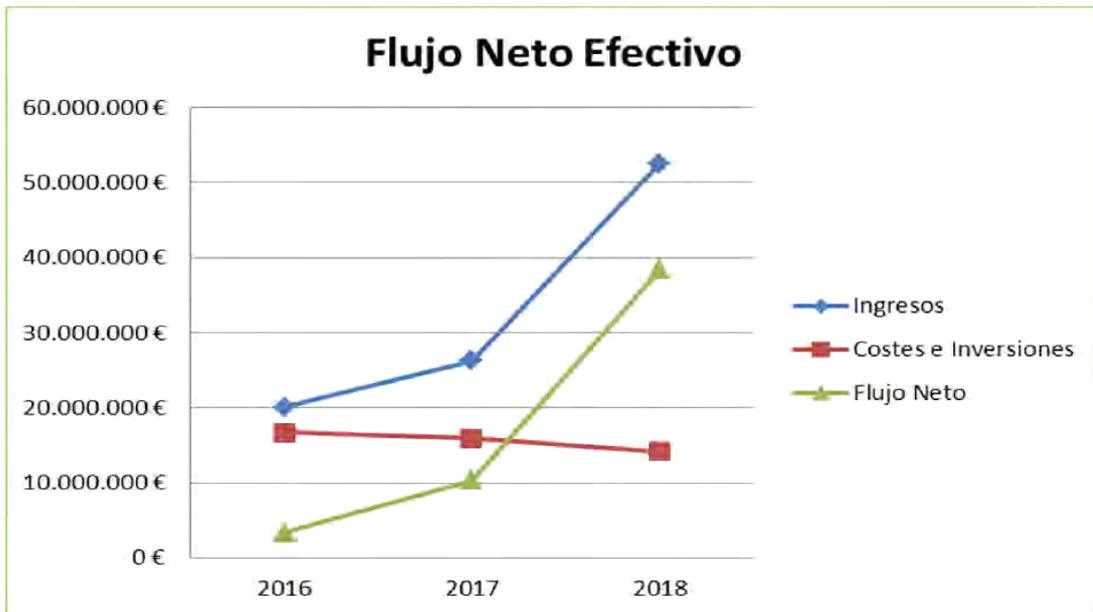


Figura 8-1. Flujo neto efectivo

Por último comentar que mediante el **análisis de sensibilidad**, es posible identificar las variaciones máximas permisibles de las magnitudes de factores tales como la inversión inicial, los flujos de caja o la tasa de interés. Los cambios en los factores pueden modificar los resultados de decisión. Con el manejo de la sensibilidad, es posible establecer los intervalos de comportamiento aceptables para que el proyecto continúe siendo rentable.

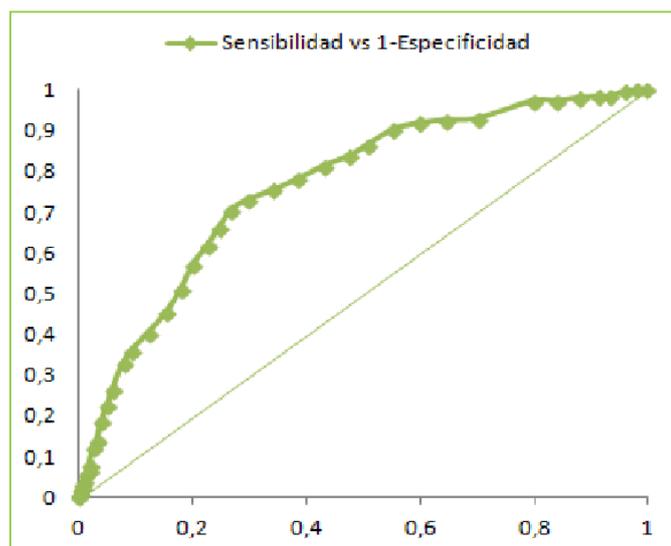


Figura 8-2. Curva ROC de reincidencia

Hay que tener en cuenta que la disminución de la tasa de reincidencia y por tanto, su rentabilidad, irá creciendo a largo plazo, ya que los resultados van en función

de las penas y de la implantación progresiva del producto en el territorio, con un periodo de adaptación, normas y revisión por parte de las Instituciones del modelo jurídico- penitenciario actual.

9 Conclusiones

El objetivo principal de la solución **Predictive Data Justice** consiste en proveer de un sistema basado en inteligencia de negocio a la **Secretaría General de Instituciones Penitencias (SGIIPP)** y entidades colaboradoras, con el objetivo de reducir el porcentaje de reclusos que cometen reincidencia delictiva, a través de un modelo analítico-predictivo que establece una serie de mejores recomendaciones basadas en posibles tratamientos (Next Best Action) personalizables y adecuadas para cada tipo de recluso para lograr su reinserción.

España tiene una tasa de reclusos de 153/100.000 habitantes, más del doble de la Unión Europea y una tasa de reincidencia delictiva superior al 50% por lo que el gasto incurrido por la **Administración General del estado (AGE)** es bastante elevado.

Predictive Data Justice proporciona una mejor adecuación de la partida presupuestaria destinadas a estas materias. Es un producto cuya implantación es económicamente viable y rentable ya que su inversión es relativamente baja con respecto al retorno que se puede llegar a obtener, **destacando el fuerte valor y dimensión social** con la mejor solución tecnológica actualmente existente, obteniendo unos porcentajes de reducción en 4 años cercanos al 16%, lo que supone un retorno de la inversión cercano al 55%, y un ahorro a las arcas públicas de más de 50 millones de euros.

10 Mejoras

En el diseño del producto **Predictive Data Justice** se han detectado nuevos puntos a desarrollar en el futuro lo que le convertirá, sin duda, en una referencia dentro del sector. Se pueden resumir en los siguientes puntos:

- ✓ **Sistema Cognitivo.**
- ✓ **Plataforma multilinguaje.**
- ✓ **Inclusión en el modelo predictivo de fuentes multimedia (video y audio).**

11 Bibliografía y referencias

Tabla 11-1. Revistas electrónicas de interés

Boletín no oficial del Área Jurídica de Pastoral Penitenciaria Católica Española
Criminológico de la Universidad de Málaga
Cuaderno del Instituto Vasco de Criminología. Eguzkilore.
Diccionario Interactivo de Derecho Penitenciario
Revista de Estudios Penales y Criminológicos de la Universidad de Santiago de Compostela
Revista de Estudios Penitenciarios. Secretaría General de Instituciones Penitenciarias
Revista Electrónica de Ciencia Penal y Criminología de la Universidad de Granada
Revista Española de Investigación Criminológica
Revista Española de Sanidad Penitenciaria
Revista Seguridad y Ciudadanía. Ministerio del Interior

Tabla 11-2. Observatorios y centros de documentación

International Centre for Prison Studies (ICPS)
Observatorio del Sistema Penal y de Derechos Humanos de la Universidad de Barcelona
Observatorio de Derecho Penitenciario del Colegio de Abogados de Pamplona
Observatorio de las Políticas y Sistemas Penitenciarios del Máster en Criminología aplicada a la ejecución de penas de la Universidad de Murcia
Observatorio de Justicia Juvenil
Observatorio Penitenciario del Departamento de Derecho Político de la UNED
Observatorio Penitenciario (Tomás Montero)

Tabla 11-3. Fuentes estadísticas

ACAIP. Dosieres y boletines estadísticos

Aumento de presos y código penal, González Sánchez, Ignacio 2011

Capdevila, Manel y Marta Ferrer Puig 2009, Tasa de reincidencia penitenciaria 2008, Generalidad de Cataluña

Centro Internacional de Estudios Penitenciarios (ICPS)

Cid, José 2008, El incremento de la población reclusa en España. Diagnóstico

Consejo de Europa. SPACE I. Statistiques pénales annuelles du Conseil de L'Europe: SPACE I y II

Departamento de Justicia. Boletines y descriptores estadísticos

Eurostat. Estadísticas de población y condiciones sociales. Delitos y justicia criminal, Generalitat de Catalunya

IDESCAT. Anuario Estadístico de Cataluña

Instituto Nacional de Estadística

Ministerio del Interior, Anuario estadístico 2007 (extracto con datos de población reclusa)

Ministerio del Interior. Anuario Estadístico

Ministerio Sanidad, Política Social e Igualdad. Instituto de la Mujer. Estadísticas

Revista Española de Investigación Criminológica, número 6

Secretaría General de Instituciones Penitenciarias. Estadísticas

12 Glosario

AGE	Administración General del Estado
APE	Administración Pública de España
ASECEDI	Asociación de Entidades de Centros de Día
CENDOJ	Centro de Documentación Judicial
CNEJ	Comité Nacional de Estadística Judicial
CPN	Cuerpo de Policía Nacional
DAFO	Debilidades, Amenazas, Fortalezas y Oportunidades
FEAPS	Confederación Española de Organizaciones en favor de las Personas con Discapacidad Intelectual
FyCSE	Fuerzas y Cuerpos de Seguridad del Estado
GC	Guardia Civil
KPI	Key Performance Indicator (Indicador clave de desempeño)
LOFAGE	Ley 6/97, del 14 de Abril, de Organización y Funcionamiento de la Administración General del Estado
LRJAP- PAC	Ley 4/1999, del 13 de Enero, de Régimen Jurídico de las Administraciones Públicas y del Procedimiento Administrativo Común
MF	Ministerio Fiscal
MI	Ministerio del Interior
MJ	Ministerio de Justicia
ONG	Organizaciones No Gubernamentales

SEC	Sistema Estadístico de Criminalidad
SGIIPP	Secretaría General de Instituciones Penitenciarias
SGNTJ	Subdirección General de Nuevas Tecnologías del Ministerio de Justicia
UNAD	Unión Española de Asociaciones y Entidades de Atención al Drogodependiente

13 Anexos

13.1 Centros Penitenciarios

La distribución de los diferentes centros penitenciarios en el territorio nacional es la que se muestra en la siguiente figura.



Figura 13-1. Centros Penitenciarios

La distribución por centros territoriales para esta primera fase de implantación incluye las siguientes regiones:

- CT8. Castilla-La Mancha y Extremadura.
- CT5. Comunidad Valenciana, Murcia e Islas Baleares.
- CT10. Sevilla, Cadiz y Huelva.

- CT8. Asturias, Cantabria y País Vasco.
- CT4. Comunidad de Madrid, únicamente los centros penitenciarios:
 - Madrid I. Mujeres. Alcalá de Henares.
 - Madrid II. Alcalá de Henares.
 - Madrid III. Valdemoro.
 - Madrid IV. Navalcarnero.

13.2 Criterios de evaluación de riesgos

En las siguientes tablas se encuentran las definiciones que se han utilizado para realizar una evaluación de los riesgos:

Entrada	Definición
Identificador	Identificador unívoco del riesgo. Este identificador estará compuesto por un contador incremental según la estructura PDJ-RIE.[XXXX].
Descripción	Breve descripción del riesgo, indicando sus posibles causas así como una descripción de los perjuicios que puede causar
Impacto	Se debe indicar el impacto del riesgo en el proyecto
Probabilidad	Se debe indicar la probabilidad de ocurrencia del riesgo en el proyecto.
Criticidad	Este campo es establecido en función del impacto y la probabilidad.
Estrategia	Refiere la principal estrategia, o combinación de estrategias para la gestión de riesgos. Los posibles valores serán: eliminar, mitigar, aceptar, transferir.
Acciones	Se indican las diversas acciones que se tomarán para su gestión.

Valoración del impacto del riesgo

Valor	Definición
Muy alto	Su ocurrencia afectará a objetivos críticos del proyecto no se cumplirán (coste, calendario, calidad o satisfacción del cliente), pudiendo incluso causar la no viabilidad del proyecto
Alto	Su ocurrencia implicará serios y significativos perjuicios en el proyecto (retrasos, aumentos de costes, modificación en el alcance) y afectará a objetivos críticos del proyecto
Medio	Su ocurrencia implicará perjuicios importantes en el proyecto
Bajo	Su ocurrencia implicará sobreesfuerzos asumibles en el proyecto

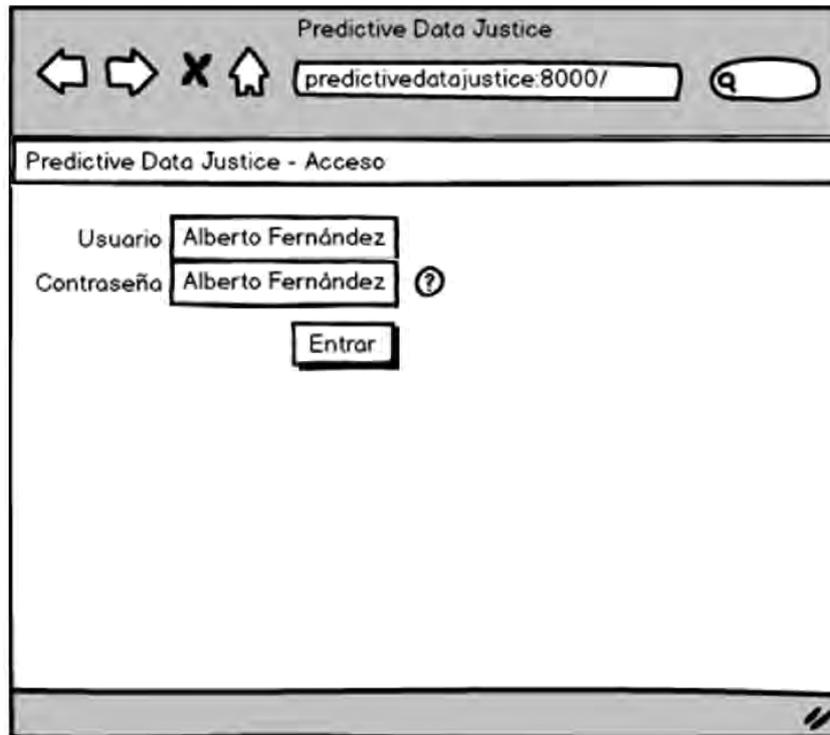
Probabilidad de ocurrencia del riesgo

Valor	Definición
Muy alta	La probabilidad es cercana a la seguridad de ocurrencia (>0.9)
Alta	La probabilidad de ocurrencia es muy elevada (0.6-0.9)
Media	La probabilidad de ocurrencia es moderada (0.3-0.6)
Baja	Difícilmente o muy difícilmente se materializará el riesgo (<0.3)

Estrategia

Valor	Definición
Evitar	Implica tomar todas las medidas necesarias para la eliminación total de la amenaza
Transferir	Implica trasladar a un tercero todo o parte del impacto negativo de una amenaza, junto con la propiedad de respuesta. Esto confiere a una tercera persona la propiedad de la responsabilidad de su gestión; pero no lo elimina
Mitigar	Implica tomar las medidas necesarias para reducir a un umbral (nivel) aceptable la probabilidad y/o el impacto del evento. Se consigue aplicando técnicas de prevención, que ayuden a reducir la probabilidad de ocurrencia del riesgo. En aquellos casos, que no pueda reducirse la probabilidad de ocurrencia, deberá de dirigir los esfuerzos en reducir la severidad del mismo, es decir, añadir contramedidas que reduzcan el impacto del mismo
Aceptar	Se adopta la técnica de aceptar los riesgos, de dos maneras: activas o pasiva. La activa corresponde a disponer de un plan de contingencias para hacer frente a la amenaza, y la pasiva en dejar que el equipo de trabajo gestione el riesgo conforme se presenta

13.3 Mock-ups de la aplicación visual



Predictive Data Justice

← → ✕ 🏠 predictivedatajustice:8000/ 🔍

Predictive Data Justice - Listado de acciones

Filtros

Organización ? FAD

Fechas ? 27/06/15 01/12/15

Coste ? min max Aportación ? min max Van a caducar ? Solo novedades ?

🔍 Texto libre

📅 FEB 2008

Nombre ▲	Org ⇅	Coste	Fechas ▼	
Curso 01	FAD	2K	22/07/15-01/11/15	○
Curso 02	FAD	1K	25/08/15-04/11/15	●
Curso 03	FAD	3K	07/09/15-07/11/15	○

Borrar Nueva Editar

Predictive Data Justice

← → ✕ 🏠 predictivedatajustice:8000/ 🔍

Predictive Data Justice - Perfil del recluso

Alberto Fernández

📅 FEB 2008

📄 ?

Notas

Personales

Penitenciarios

Judiciales

Análisis

Acciones

Datos Personales de Alberto Fernández

Edad: 45 años

Estudios superiores en Big Data

Casado

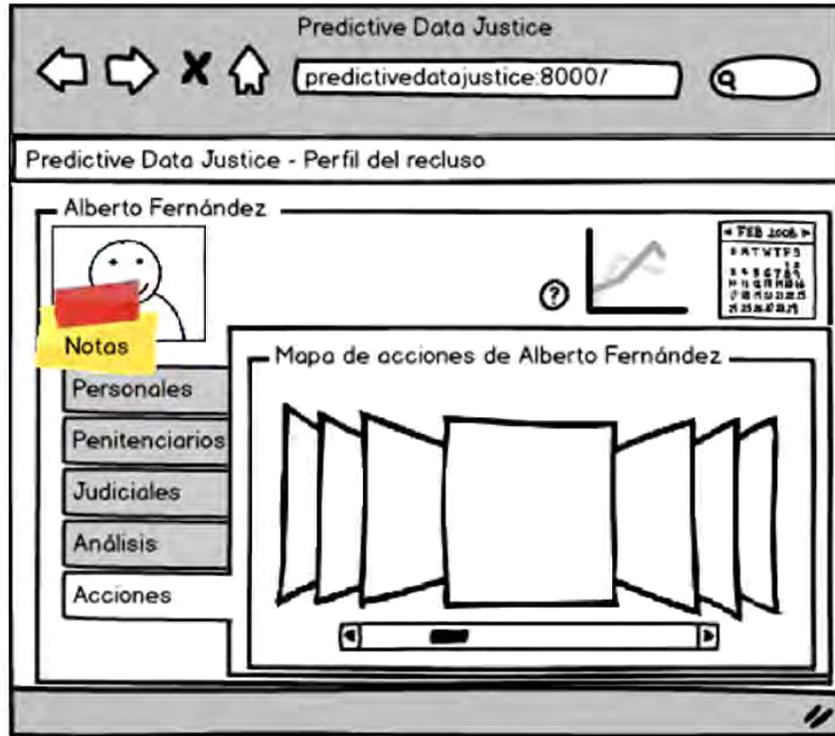
No presenta discapacidades

0% de minusvalía

No consumo de drogas

Salud estable





13.4 Código fuente en R

```
#####  
#####  
  
#NUBE DE TAGS  
  
#####  
#####  
  
library(tm)  
  
options(max.print=5.5E5)  
  
txt <- system.file("texts", "txt", package = "tm")  
  
txt  
  
class(txt)  
  
ini_corpus <- Corpus(DirSource(txt), readerControl =  
list(language="lat"))  
  
corpus <- ini_corpus  
  
summary(corpus)  
  
inspect(corpus)  
  
corpus <- tm_map(corpus, removeNumbers)  
corpus <- tm_map(corpus, removePunctuation)  
corpus <- tm_map(corpus, stripWhitespace)  
corpus <- tm_map(corpus, removeWords, stopwords("spanish"))  
  
a <- DocumentTermMatrix(corpus)  
a <- removeSparseTerms(a, 0.9)  
  
#a  
  
#inspect(a)  
  
sink(file="C:\\s/output.txt")  
  
inspect(a)  
  
sink(NULL)  
  
library(wordcloud)  
  
library(Rstem)  
  
m <- as.matrix(a)
```

```
#m
word_list <- sort(colnames(m), decreasing = TRUE)
#word_list
d <- data.frame(freq = sort(colSums(m), decreasing = TRUE))
#d
d$stem <- wordStem(row.names(d), language = "spanish")
#d
d$word <- row.names(d)
# remove web address (very long string):
d <- d[nchar(row.names(d)) < 20, ]
d <- d[nchar(row.names(d)) > 6, ]
d
agg_freq <- aggregate(freq ~ stem, data = d, sum)
agg_word <- aggregate(word ~ stem, data = d, function(x) x[1])
d <- cbind(freq = agg_freq[, 2], agg_word)
d <- d[order(d$freq, decreasing = T), ]
wordcloud(d$word, d$freq, random.order=FALSE, colors=brewer.pal(8,
"Dark2"))
# Pintando Dendograma
x <- as.data.frame(inspect(m))
```

```
#####  
#####  
  
#DENDROGRAMA  
  
#####  
#####  
  
library(twitter)  
  
require(tm)  
  
library(wordcloud)  
  
library(RColorBrewer)  
  
library(RCurl)  
  
options(max.print=5.5E5)  
  
txt <- system.file("texts", "txt", package = "tm")  
  
txt  
  
class(txt)  
  
ini_corpus <- Corpus(DirSource(txt), readerControl =  
list(language="lat"))  
  
corpus <- ini_corpus  
  
summary(corpus)  
  
#inspect(corpus)  
  
corpus <- tm_map(corpus, removeNumbers)  
corpus <- tm_map(corpus, removePunctuation)  
corpus <- tm_map(corpus, stripWhitespace)  
corpus <- tm_map(corpus, removeWords, stopwords("spanish"))  
#corpus <- tm_map(corpus, stemDocument, language = "spanish")  
mydata.dtm = TermDocumentMatrix(corpus)  
mydata.dtm2 = DocumentTermMatrix(corpus)  
mydata.dtm  
  
# inspeccionamos las palabras más relevantes  
#findFreqTerms(mydata.dtm, lowfreq=10, highfreq = 15)
```

```
mydata.dtm2 = removeSparseTerms(mydata.dtm, sparse=.3)
mydata.dtm2
# Transformamos la matriz en un data frame
mydata.df = as.data.frame(inspect(mydata.dtm2))
# hacemos algunas comprobaciones sobre el dataframe
nrow(mydata.df)
ncol(mydata.df)
mydata.df.scale = scale(mydata.df)
d = dist(mydata.df.scale, method = "euclidean") # definimos la
distancia
fit = hclust(d, method="ward.D") # definimos el método para hacer
clusters
## Dibujamos el dendograma
# Generamos el dendograma
plot(fit, main="")
# Ponemos un título al dendograma
titulo = c("Dendograma en Predictive Data Justice ")
title(titulo)
## Parte 2
## Dibujamos los clusters
k = "10"
k = as.numeric(unlist(strsplit(k, ",")))
rect.hclust(fit, k, border=rainbow(k))
```

```
#####
#####
```

```
#RELACIONES

#####

#####

library(twitter)

require(tm)

library(wordcloud)

library(RColorBrewer)

library(RCurl)

options(max.print=5.5E5)

txt <- system.file("texts", "txt", package = "tm")

txt

class(txt)

ini_corpus <- Corpus(DirSource(txt), readerControl =
list(language="lat"))

corpus <- ini_corpus

summary(corpus)

#inspect(corpus)

corpus <- tm_map(corpus, removeNumbers)

corpus <- tm_map(corpus, removePunctuation)

corpus <- tm_map(corpus, stripWhitespace)

corpus <- tm_map(corpus, removeWords, stopwords("spanish"))

#corpus <- tm_map(corpus, stemDocument, language = "spanish")

mydata.dtm = TermDocumentMatrix(corpus)

#mydata.dtm2 = DocumentTermMatrix(corpus)

mydata.dtm

# inspeccionamos las palabras más relevantes

#findFreqTerms(mydata.dtm, lowfreq=10, highfreq = 15)

#mydata.dtm2 = removeSparseTerms(mydata.dtm, sparse=.9)

#mydata.dtm2

freq <- colSums(as.matrix(mydata.dtm))
```

```
findFreqTerms(mydata.dtm, lowfreq=1, highfreq = 50)
# encontrar relaciones de palabras especificas
#findAssocs(mydata.dtm2, "antecedent", corlimit=0.7)
f <- findAssocs(mydata.dtm, "antecedentes", corlimit=0.58)
f
#findAssocs(mydata.dtm2, "condena", corlimit=0.6)
library(Rgraphviz)
##Dibujamos las relaciones
#plot(mydata.dtm2, terms=findFreqTerms(mydata.dtm2,
lowfreq=20)[1:6], corThreshold=0.2)
#plot(mydata.dtm2)
#plot(mydata.dtm2, terms=f, corThreshold=0.5)
```

```
#####
#####
#KMEANS
#####
#####
```

```
library(twitterR)

require(tm)

library(wordcloud)

library(RColorBrewer)

library(RCurl)

options(max.print=5.5E5)

txt <- system.file("texts", "txt", package = "tm")

txt

class(txt)

ini_corpus <- Corpus(DirSource(txt), readerControl =
list(language="lat"))

corpus <- ini_corpus

summary(corpus)

#inspect(corpus)

corpus <- tm_map(corpus, removeNumbers)

corpus <- tm_map(corpus, removePunctuation)

corpus <- tm_map(corpus, stripWhitespace)

corpus <- tm_map(corpus, removeWords, stopwords("spanish"))

#corpus <- tm_map(corpus, stemDocument, language = "spanish")

mydata.dtm = TermDocumentMatrix(corpus)

mydata.dtm2 = DocumentTermMatrix(corpus)

plot(mydata.dtm, terms=findFreqTerms(mydata.dtm,
lowfreq=100)[1:50], corThreshold=0.6)

dtm <- mydata.dtm

library("RSiteCatalyst")

library("RTextTools") #Loads many packages useful for text mining

searchkeywords <- QueueRanked(<report_suite>, "2013-02-01", "2013-
09-16",

                                c("entries", "visits", "pageviews",
"instances", "bounces"),
```

```
                                "searchenginaturalkeyword",
top="100000", startingWith = "1")

dtm <- create_matrix(searchkeywords$'Natural Search Keyword',
                    stemWords=TRUE,
                    removeStopwords=FALSE,
                    minWordLength=1,
                    removePunctuation= TRUE)

searchkeywords <- c("antecedentes", "sexo", "pene", "droga",
"marihuana", "cocaina", "reincidencia", "delito")

kmeans5<- kmeans(dtm, 5)

kmeans5

kw_with_cluster <- as.data.frame(cbind(searchkeywords$'Natural
Search Keyword', kmeans5$cluster))

names(kw_with_cluster) <- c("keyword", "kmeans5")

cluster1 <- subset(kw_with_cluster, subset=kmeans5 == 1)
cluster2 <- subset(kw_with_cluster, subset=kmeans5 == 2)
cluster3 <- subset(kw_with_cluster, subset=kmeans5 == 3)
cluster4 <- subset(kw_with_cluster, subset=kmeans5 == 4)
cluster5 <- subset(kw_with_cluster, subset=kmeans5 == 5)
```


Índice

1	Problema de negocio	2
1.1	Situación actual.....	2
1.2	El reto: Objetivo y oportunidad.....	3
1.3	Motivación y viabilidad	5
2	¿Quiénes son nuestros clientes?	5
2.1	Descripción del marco cliente.....	5
2.2	Otros posibles clientes.....	6
3	¿Quiénes somos?	6
3.1	Modelo de negocio.....	7
4	Solución	9
4.1	Fuentes de datos.....	9
4.2	Arquitectura	10
4.3	Modelo predictivo.....	11
4.3.1	Objetivos	11
4.3.2	Planteamiento y finalidad.....	11
4.3.3	Definición de variables.....	11
4.3.4	Muestra	11
4.3.5	Instrumentos utilizados.....	12
4.3.6	Fases	12
4.3.7	Medida de clasificación del modelo.....	13
4.4	Motor de recomendaciones.....	13
4.5	Funcionalidades	13
5	Plan de implantación.....	14
5.1	Riesgos	15
6	Business Case	15
6.1	Estructura de costes.....	16
6.2	Rentabilidad financiera.....	17
7	Conclusiones.....	18
8	Mejoras.....	18
9	Glosario de términos.....	19

1 Problema de negocio

Existe la creencia generalizada de que los delincuentes que se reinserían en la sociedad, presentan una alta probabilidad de que vuelvan a cometer un delito. Sin embargo, la distribución de la reincidencia delictiva, es en realidad muy heterogénea y varía en función del delito cometido.

Consecuentemente, es evidente la importancia que tiene analizar los factores que están implicados en el riesgo de reincidencia delictiva. Bajo este contexto nace nuestro proyecto de **Predictive Data Justice**, una solución para la **evaluación del riesgo de reincidencia delictiva entre los delincuentes**, mediante modelos predictivos, analizando de forma conveniente la influencia de todos los posibles factores que confluyan en cada caso, para **establecer una serie de recomendaciones basadas en posibles tratamientos**, que logren una completa rehabilitación del delincuente.

La solución sirve para mejorar la aplicabilidad de las **políticas de prevención delictiva**, optimizar el proceso penitenciario, con una mejor asignación de las actividades terapéutico-asistenciales, directamente dirigidas y orientadas hacia la **consecución efectiva de la reeducación y reinserción social del delincuente**, y aportar un gran valor social para la ciudadanía, incrementando el estado de bienestar y la seguridad ciudadana.

1.1 Situación actual

Desde un punto de vista estadístico, **España lidera el ranking de los países de Europa Occidental en cuanto a número de reclusos y centros penitenciarios se refiere**. Los datos referidos en este capítulo son extraídos del informe ODA (Realidad y Política penitenciarias) así como el boletín criminológico ACAIP.

España posee una tasa de reclusos situada en 153 por cada 100.000 habitantes, más del doble de la Unión Europea, situada en 70. Contrasta fuertemente con otras tasas como los 84,6 de Alemania, los 111,6 de Italia y los 117 por cada 100.000 habitantes en Francia. Paradójicamente España es uno de los países con cifras de criminalidad más baja de Europa: 46,1 delitos por cada mil habitantes, frente a los 62,8 de media europea.

Tabla 1-1. Estadísticas penitenciarias de España frente a Europa

	España	Europa
Tiempo medio de estancia en centros penitenciarios	19 meses	10 meses
Porcentaje de reclusos por robo	30,70%	12,90%
Tasa de reclusos por 100.000 habitantes	153/100.000	70/100.000
Posición en el ranking de número de reclusos	4º Posición	
Posición en el ranking de condenas más extensas	6ª Posición	

El **Consejo Europeo** urge a adoptar medidas, entre las que destacan la incentivación de políticas de reeducación y de reinserción social, como forma principal y alternativa a la privación de la libertad.

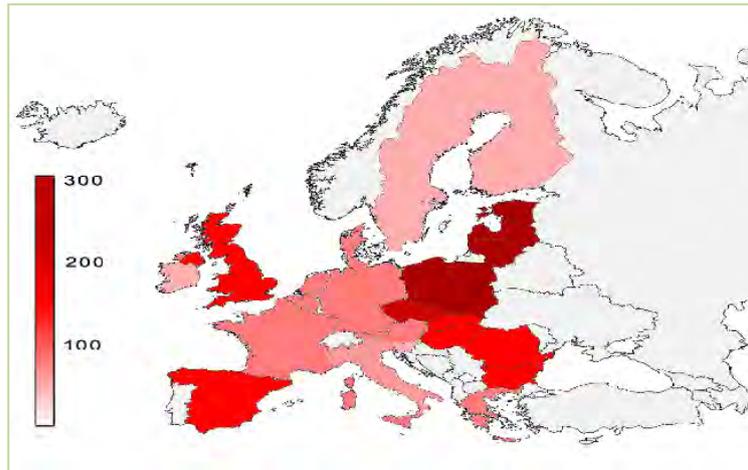


Figura 1-1. Tasa de población carcelaria por cada 100.000 habitantes

Según un estudio de la Fundación Atenea, **dos de cada tres reclusos vuelve a cometer un delito cuando queda puesto en libertad, lo que revela una elevada tasa de reincidencia entre los reclusos españoles.**

Por otra parte se calcula que el 80% de los reclusos consumían drogas antes de ingresar en un centro penitenciario. Este dato puede explicar el hecho de que **España**, pese a ser uno de los países con menor delincuencia, **se encuentra entre los primeros países de Europa por delitos contra la salud pública**, por delante de Francia, Italia y Alemania.

El **Consejo Europeo**, alarma sobre la masificación de los centros penitenciarios en países como España y Reino Unido, y **urge a utilizar vías de castigo alternativas.**

1.2 El reto: Objetivo y oportunidad

El análisis de los registros criminales es fundamental en la prevención del delito. En la **Administración General del Estado (AGE)**, este tipo de análisis se ha realizado **históricamente mediante herramientas estadísticas descriptivas básicas**, que no refleja la verdadera interrelación de las variables y por lo tanto, no aporta una solución efectiva al problema real.

En los últimos años, tanto **el Ministerio del Interior (MIR) como el Ministerio de Justicia (MJ)** han liderado un importante avance tecnológico en busca de un claro objetivo, la digitalización de todos sus expedientes. Gracias a la irrupción de este **nuevo paradigma digital** surgen nuevas oportunidades de negocio. La oportunidad de aprovechar un entorno completamente digitalizado, unido al tratamiento masivo de la información existente que proviene de diversas fuentes a priori no interconectadas entre sí, abre una nueva puerta de mercado al mundo de la inteligencia de negocio en la toma de decisiones aplicable dentro del sistema penitenciario, que permita una **evaluación del riesgo de reincidencia delictiva.**

El gasto incurrido por la AGE en materias de penitenciaría, seguridad ciudadana y justicia, se encuentra dentro de los Presupuestos Generales del Estado (PGE), como un gasto básico y de carácter obligatorio. El objetivo consiste en **proporcionar una mejor adecuación de la partida presupuestaria destinada a la SGIIPP**, mediante la optimización del proceso penitenciario, directamente orientado hacia la consecución efectiva de la reinserción social del delincuente.



Figura 1-2. Impacto sobre los procesos judiciales y de reinserción en un individuo

Oportunidades del proyecto

- ✓ La información procesal se encuentra digitalizada como dato no estructurado.
- ✓ Las herramientas actuales no permiten extraer conclusiones reveladoras del problema.
- ✓ Falta de competidores en el ámbito Big Data.
- ✓ Minería de texto sin ironía.

Retos del proyecto

- ✓ Implantar un potente análisis de datos dentro del proceso penitenciario.
- ✓ Focalizar en el caso de uso de reincidencia delictiva.
- ✓ Transmitir el conocimiento de las nuevas tecnologías basadas en Big Data.
- ✓ Abrir un nuevo modelo de negocio altamente innovador que aporte un gran valor a la AGE.

Objetivos del proyecto

- ✓ Descongestión de los centros penitenciarios.
- ✓ Reducción de la tasa de reclusos reincidentes (pasando del 55% actual al 28.5% en el año 2018).
- ✓ Reducción del tiempo medio de estancia en los centros penitenciarios (pasando de los 19 meses actuales, al valor actual de la media de la Unión Europea en el año 2018).
- ✓ Transformación del proceso penitenciario.

1.3 Motivación y viabilidad

¿Por qué Predictive Data Justice es un proyecto innovador?



- ✓ Aporta una dimensión social.
- ✓ Nuevas tecnologías en el entorno penitenciario.
- ✓ Aprovechamiento del dato procesal y penitenciario.
- ✓ Minería de texto aplicada a un entorno jurídico-penitenciario completamente veraz, con una completa ausencia de la ironía.

¿Qué aporta el Big Data a la solución Predictive Data Justice?



- ✓ Extraer conocimiento y valor.
- ✓ Desarrollo y construcción de una solución económica, con una gran capacidad de almacenamiento y proceso.
- ✓ Solución altamente escalable.
- ✓ Analizar textos previamente digitalizados sin necesidad de estructurarlos ni de almacenarlos en una base de datos relacional.

¿Por qué Predictive Data Justice es un proyecto económicamente viable?



- ✓ La inversión inicial es baja con respecto al retorno que se puede llegar a obtener.
- ✓ Bajo coste de mantenimiento.
- ✓ La AGE está apostando por la inversión en este tipo paradigma tecnológico del Big Data.

2 ¿Quiénes son nuestros clientes?

2.1 Descripción del marco cliente

Este proyecto queda enmarcado, de una forma general, en el ámbito de la **Administración General del Estado (AGE)**.

El cliente fundamental y principal al que va dirigido este proyecto es la **Secretaría General de Instituciones Penitenciarias (SGIIPP)**, dependiente de la Secretaría de Estado de Seguridad del Ministerio del Interior (MIR), cuyos fines son garantizar el cumplimiento de las penas impuestas por los jueces, en la reeducación y reinserción social de los sentenciados a penas privativas de libertad.

Así mismo, consideramos intervinientes clave de nuestro proyecto, todas aquellas entidades colaboradoras con la SGIIPP, como pueden ser las **Organizaciones No Gubernamentales y Fundaciones** (Unión Española de Asociaciones y Entidades de Atención al Drogodependiente, Cruz Roja Española, Proyecto Hombre, Fundación Atenea, etc.), cuyo papel resulta esencial para la colaboración y participación de la sociedad a la hora de conseguir la reinserción y reeducación de las personas condenadas.

Las **Fuerzas y Cuerpos de Seguridad del Estado (FyCSE)** se consideran como potenciales clientes de este proyecto, ya que están bajo la dependencia de la Secretaría de Estado de

Seguridad del Ministerio del Interior (MI), Por otra parte, es necesario incorporar también como cliente de nuestro proyecto al **Ministerio de Justicia (MJ)**, departamento ministerial que como legislador se encarga de ejecutar el desarrollo del ordenamiento jurídico.

La siguiente figura representa las relaciones más importantes que se establecen entre nuestro principal cliente (la SGIIPP), y sus entidades colaboradoras.



Figura 2-1. Relaciones entre la SGIIPP y sus entidades colaboradoras.

2.2 Otros posibles clientes

Nuestro proyecto **Predictive Data Justice** es perfectamente escalable a otros clientes del **panorama internacional**. Se trata de una nueva oportunidad de negocio, que, inicialmente, se podrá implantar sin necesidad de modificarla en otros países que compartan ese mismo idioma.

Del mismo modo, otros posibles interesados en nuestro proyecto, son los **abogados**, constituidos o no en bufetes. La idea reside en la oportunidad de proporcionarles una nueva forma de obtener información estadística sobre futuros procesos judiciales, teniendo en cuenta los perfiles de los implicados en los mismos. De esta forma, los abogados podrán predecir el porcentaje de éxito de sus futuros procesos judiciales.

3 ¿Quiénes somos?

Predictive Data, S.L. es una empresa especializada en el diseño, desarrollo e integración de soluciones **Big Data & Business Analytics** dentro del **ámbito público internacional** (Administraciones, Gobiernos, Ministerios, etc.), extrayendo automáticamente información de todo tipo de contenidos para darle valor, aplicando tecnologías semánticas (minería de texto) y un potente motor de recomendaciones.

Tenemos distintas soluciones o productos en el mercado internacional para los sectores de Sanidad, Transporte y Servicios de carácter público.

En el ámbito jurídico-penitenciario, se plantean nuevos retos y requerimientos que los actuales sistemas de gestión documental no son capaces de soportar.

Dada la alta relevancia de los contenidos presentes en la documentación jurídica y penitenciaria, resulta imprescindible identificar y extraer, de forma automática, determinados conceptos que se manejan, utilizando para ello unos modelos que denominamos como ontologías.

Estos conceptos son automáticamente reconocidos y extraídos de los contenidos documentales, gracias a la solución analítica que se propone en **Predictive Data Justice**.

A través de **Predictive Data Justice**, la asignación de recursos humanos y materiales - optimizando el gasto presupuestario-, las situaciones de prisión provisional y la conveniencia de conceder o no un determinado beneficio penitenciario, estarán mejor apoyados por nuestra solución, que se alimenta de grandes bases de datos y opera sobre potentes y novedosos modelos y algoritmos previamente definidos.

3.1 Modelo de negocio

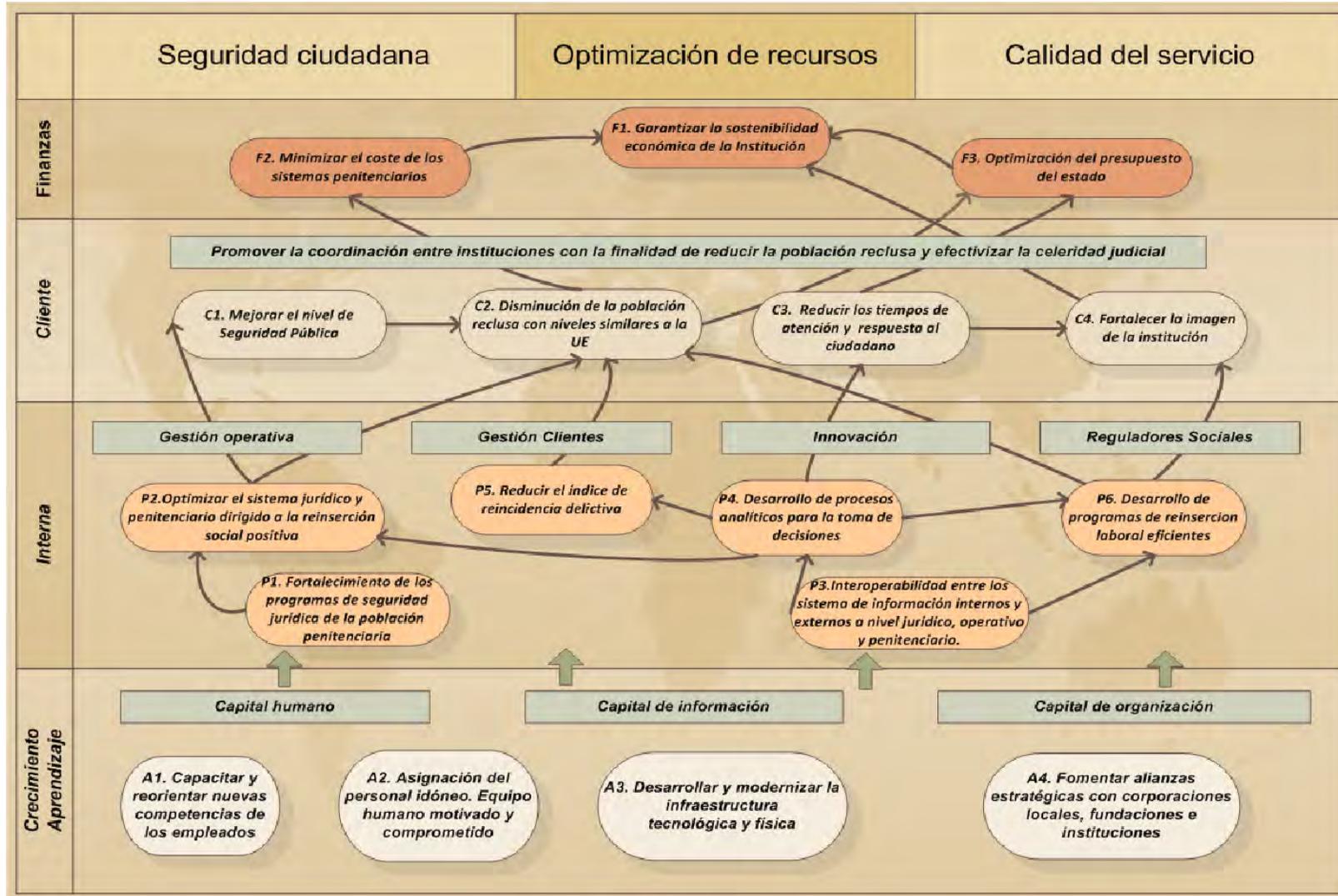


Figura 3-1. Modelo de negocio

Nuestra **propuesta de valor** es optimización de la asignación de las acciones de reinserción con el objetivo de minimizar la reincidencia delictiva.

Cabe destacar de este proyecto su fuerte **valor social**, que garantiza un **servicio público en el ámbito de la Justicia y la seguridad ciudadana** eficientes y de gran calidad, tal y como exigen tanto la ciudadanía como la población penitenciaria a través de sus derechos reconocidos, beneficiándose de todo ello el conjunto de la sociedad.

A continuación se presenta el mapa estratégico de la empresa.



4 Solución

Predictive Data Justice se basa en la captura de **datos internos y externos, estructurados y no estructurados**, tratado con un **modelo predictivo** previamente construido y entrenado a partir de los datos socio demográficos, penales, penitenciarios, familiares, escolares, laborales y de salud mental de los reclusos para finalmente recomendar (**Next Best Action**), de una forma específica a cada recluso que presente un elevado riesgo de cometer reincidencia delictiva, posibles tratamientos (dentro del catálogo de servicios proporcionados por las entidades que colaboran con la SGIIP) para lograr una completa rehabilitación del recluso.

Para el almacenamiento y procesamiento de datos no estructurados provenientes fundamentalmente de redes sociales e Internet, así como de documentos previamente digitalizados (atestados y denuncias policiales, sentencias y autos judiciales, diligencias, y demás informes etc.), la solución se apoya en herramientas “Open Source”, que se encuentran disponibles en el ecosistema **Apache Spark**.

4.1 Fuentes de datos

Para alimentar la solución propuesta, se utilizan las siguientes fuentes de datos:

- ✓ **Expediente Judicial:** Sentencias, atestados y autos obtenidos del CENDOJ y el EJE (Expediente Judicial Electrónico).
- ✓ **Expediente Penitenciario:** Fichero de datos de carácter personal del Sistema Informático Social Penitenciario (SISPE) y los informes generados por los educadores, psicólogos, sociólogos, pedagogos y trabajadores sociales del centro.
- ✓ **Redes sociales e Internet:** Contenido publicado en Internet con contenidos altamente relevantes con datos referentes al entorno que le rodea, posibles adicciones e impresiones acerca de su paso por prisión para evaluar la efectividad de las acciones.



Figura 4-1. Proceso del reincidente y fuentes de información

4.2 Arquitectura

Partiendo de las fuentes de datos se construye el Data Lake sobre el cual se asienta la solución. El acceso a dicho Data Lake está securizado, mediante un módulo de gestión de usuarios capaz de controlar los accesos. Todas las posibles acciones que pueden realizar los usuarios, quedan registradas en ficheros "log" que son almacenados en el Data Lake.

Predictive Data Justice se nutre de herramientas Software del tipo "Open Source", ya que combinadas entre sí, son capaces de brindar una alta capacidad de rendimiento y potencia de procesamiento, con una gran escalabilidad, flexibilidad, y modularidad semejantes a las herramientas comercializadas bajo licenciamiento.

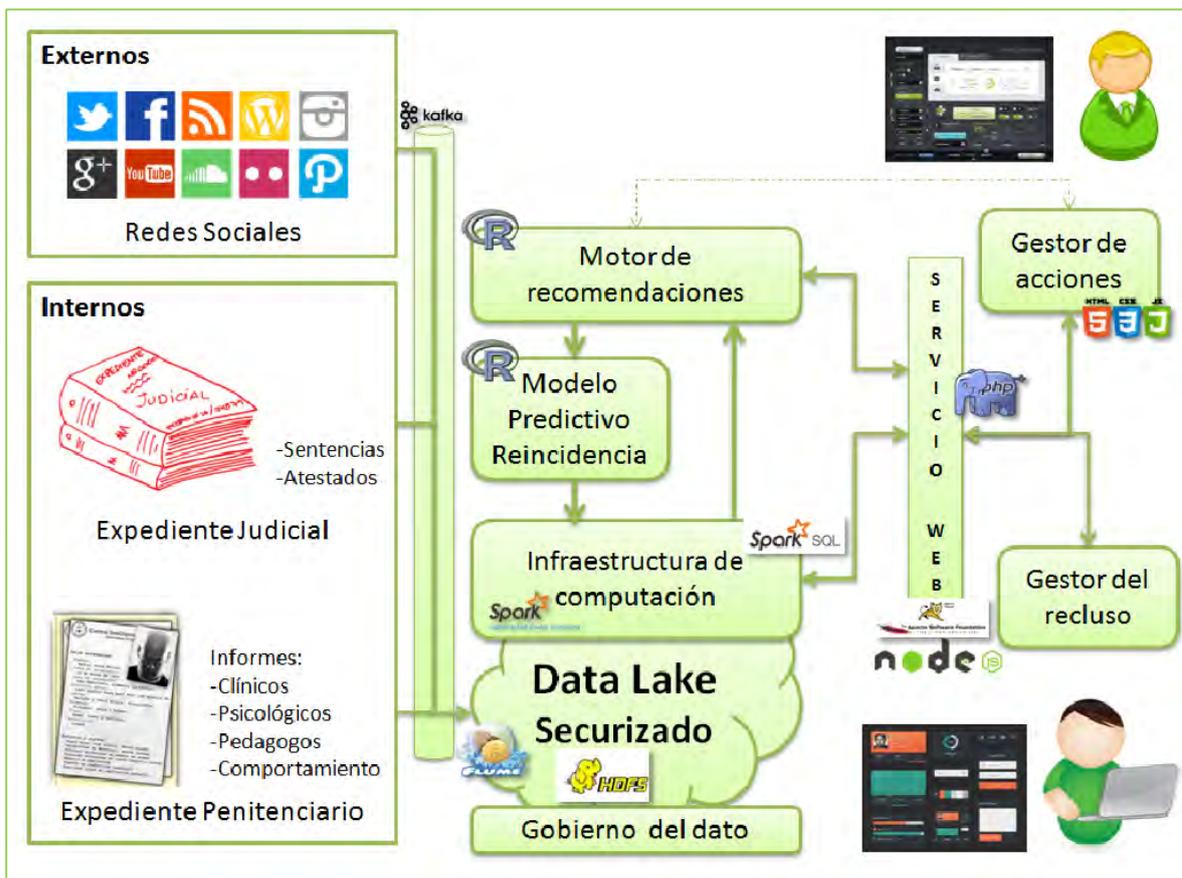


Figura 4-2. Arquitectura. Stack tecnológico de Predictive Data Justice

Para evaluar el riesgo de reincidencia delictiva de un determinado recluso, el usuario del sistema envía una petición al sistema, que tras la ejecución de un **modelo matemático analítico-predictivo**, y combinado el resultado obtenido con la acción de un **motor de recomendaciones**, podrá visualizar el correspondiente riesgo de reincidencia delictiva, junto con un listado de acciones propuestas recomendadas para aplicar sobre el recluso, con el fin de lograr reducir el riesgo de reincidencia delictiva del recluso. Todas esas acciones se introducen en el motor de recomendaciones a través del gestor de acciones.

Los modelos predictivos de reincidencia así como el modelo del motor de recomendaciones son definidos **en lenguaje R** y ejecutados sobre **SparkR**.

Los resultados de consultas, logs así como los catálogos estructurados relativos a los reclusos son almacenados igualmente en el Data Lake mediante **SparkSQL** que ofrece un acceso rápido vía WEB.

El producto **Predictive Data Justice** incorpora por defecto módulos de escalabilidad de la solución como por ejemplo Node.js y la propia plataforma Spark. El cliente no tiene que preocuparse por cómo va a crecer la aplicación en datos o en usuarios, Predictive Data Justice lo soportará a nivel de software.

El acceso para cada uno de los usuarios finales de la solución que se propone en el proyecto **Predictive Data Justice**, se realiza mediante un **portal Web** construido también íntegramente con herramientas del tipo “Open Source”. El propósito final de dicho portal es proporcionar unas potentes herramientas de Inteligencia de Negocio para los distintos perfiles que acceden a **Predictive Data Justice**. Dichos perfiles quedan estructurados en los siguientes grupos: educadores y pedagogos, sociólogos, psicólogos y resto de personal (trabajadores sociales, monitores, etc.).

4.3 Modelo predictivo

4.3.1 Objetivos

- ✓ Identificar los factores asociados con la reincidencia delictiva.
- ✓ Comprobar la capacidad de los factores para predecir la reincidencia delictiva.
- ✓ Obtener el riesgo de la reincidencia delictiva por recluso.

4.3.2 Planteamiento y finalidad

El modelo predictivo se basa en estudios realizados de forma longitudinal sobre un grupo de control, con el fin de realizar un seguimiento de los reclusos para evaluar de forma prospectiva el riesgo de reincidencia.

4.3.3 Definición de variables

Se consideran como variables básicas del estudio (o variables dependientes) el **comportamiento delictivo violento** contra las personas (homicidio y sus formas, lesiones, agresión sexual robo con violencia o intimidación), y la **reincidencia delictiva**, siendo reincidente aquel individuo que cometió delitos en el pasado, con un número de ingresos en prisión igual o superior a 2.

4.3.4 Muestra

Se utiliza un tipo de **muestreo estratificado** con vistas a aumentar la eficiencia (mayor precisión a igualdad de tamaño), sobre 811 reclusos que acumulan 953 delitos.

4.3.5 Instrumentos utilizados

- **Historial Criminológico y Social (HCS):** Sirve para disponer de información relativa a múltiples aspectos considerados como factores asociados a la reincidencia delictiva y al comportamiento delictivo violento (datos penales, penitenciarios, familiares, escolares, del nivel socioeconómico, de los grupos de relación, ocio y tiempo libre, de conductas adictivas, de consumo de drogas y alcohol, y de salud mental).
- **Inventario de Factores de Riesgo e Intervención en Prisiones (IFRIP):** Sirve para evaluar los factores de riesgo para la reincidencia delictiva y establecer los clústeres que conforman indicadores de riesgo de reincidencia, englobando **factores de riesgo estáticos** y **factores de riesgo dinámicos** o necesidades criminógenas (Historial delictivo, educación formal/empleo, financiero, familiar/ marital, vivienda/alojamiento, ocio/diversión, red social, abuso de alcohol/drogas, salud mental y actitudes/valores).

4.3.6 Fases

4.3.6.1 Análisis descriptivo de la muestra en función del HCS

Se obtienen los **estadísticos descriptivos** (aspectos demográficos y sociales relacionado con el delito) en relación con cada una de las variables que conforman el Historial Criminológico y Social (HCS).

4.3.6.2 Análisis de la reincidencia delictiva

Se denomina **reincidente** a aquel recluso que cometió en el pasado uno o más delitos y por los que cumplió condena en un centro penitenciario. Por tanto, el recluso cuenta con dos o más delitos en su carrera delictiva y reingresó en un centro penitenciario.

Se realiza un **estudio bivariado** para analizar la reincidencia delictiva en función de algunas de las variables contenidas en el Historial Criminológico y Social (HCS).

4.3.6.3 Análisis del IFRIP

Se realiza el **análisis estadístico** del Inventario de Factores de Riesgo e Intervención en Prisiones (IFRIP), con especial atención a sus propiedades psicométricas y su validez predictiva, y se construye la **curva ROC** para analizar su capacidad predictiva sobre la reincidencia delictiva, asumiendo un **área bajo la curva** (AUC) como predictor continuo.

Se define la AUC como la probabilidad que un recluso seleccionado aleatoriamente con alto riesgo específico (sobre el IFRIP) tiene de reincidir, en comparación con otro recluso, también seleccionado aleatoriamente, con bajo riesgo.

4.3.6.4 Modelo explicativo de la reincidencia delictiva

Se emplean los **modelos de ecuaciones estructurales** (AMOS) para evaluar la interrelación entre los factores del IFRIP y la reincidencia delictiva, por un lado, y entre estos factores y la violencia delictiva por otro. Consideramos como variables dependientes la reincidencia y la violencia delictiva entendidas tal como anteriormente explicamos.

Se utilizan técnicas de **análisis multivariante** para construir un modelo de propósito predictivo y para medir el efecto de una exposición en presencia de variables de confusión y de interacción.

Finalmente, **Predictive Data Justice** utiliza un **Árbol de Decisión** para establecer el riesgo de reincidencia delictiva (individualizado por recluso), cuyo conjunto de reglas se corresponde con los propios factores IFRIP.

4.3.7 Medida de clasificación del modelo

En realidad, el valor del AUC obtenido corresponde a una puntuación del IFRIP que mejor permite clasificar a los reclusos reincidentes. Por tanto, se busca una puntuación que no comprometa la sensibilidad del instrumento así como a su especificidad, y que consecuentemente, detecte correctamente a los reclusos que reincidan pero que al mismo tiempo no de falsos positivos. En este estudio se ha estimado que el mejor punto de corte para el IFRIP es de 0,31 ya que clasifica correctamente a un **75% de los reclusos reincidentes**, ofreciendo un **34% de falsos positivos**.

4.4 Motor de recomendaciones

La parte fundamental y “Know-how” de **Predictive Data Justice**, es su motor de recomendaciones. Consiste en un conjunto de metodologías y algoritmos que permiten afrontar los procesos de personalización del recluso estableciendo un ranking con acciones propuestas que consigan reducir su riesgo de reincidencia delictiva. Dichas acciones forman en su conjunto un completo catálogo denominado “gestor de acciones”.

El motor de recomendaciones se nutre de todos los resultados que se desprenden del modelo analítico-descriptivo de evaluación para el riesgo de reincidencia delictiva. Dichos resultados son tratados por un modelo que analiza la información disponible sobre cada recluso, obteniendo una lista ordenada y personalizada de posibles acciones. En base a nuestra experiencia previa, el modelo implantado en el motor de recomendaciones se basa en la regresión logística, bastante extendida en problemas relacionados con la evaluación del riesgo.

4.5 Funcionalidades

Entre las funcionalidades más importantes de **Predictive Data Justice** destacan:

- ✓ Obtención de un scoring del riesgo de reincidencia delictiva por recluso.
- ✓ Creación de un mapa de acciones recomendadas para el recluso.
- ✓ Seguimiento del estado de valor del recluso relativo a su nivel de reincidencia.
- ✓ Evaluación de cada una de las acciones propuestas por el motor de recomendaciones, con el fin de mejorar la toma de decisiones a la hora de asignarlas a los reclusos.
- ✓ Seguimiento del grado de éxito de las acciones.

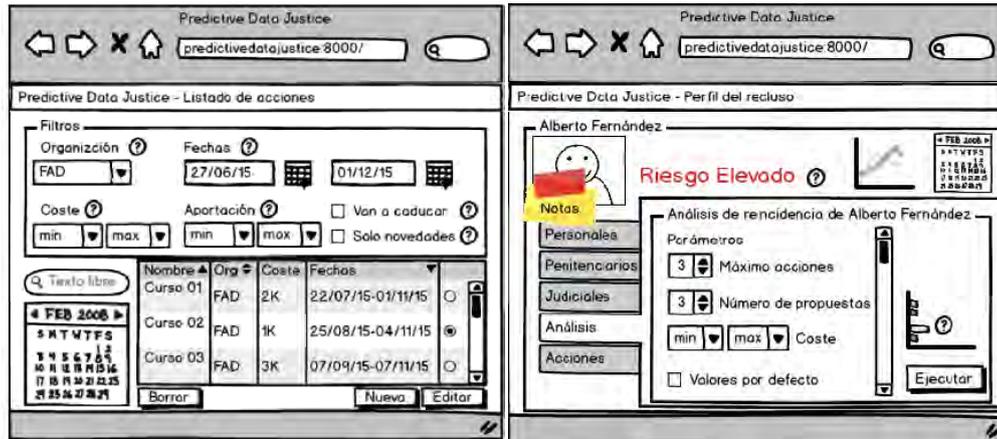


Figura 4-6. Visualización de Predictive Data Justice

5 Plan de implantación

Para la implantación de Predictive Data Justice es recomendable adaptar el estándar internacional de **metodología para Gestión y Dirección de Proyectos PMP**.

Teniendo en cuenta la distribución de las **10 Coordinaciones Territoriales (CT)** de trabajos penitenciarios en el territorio español y el informe estadístico de la población reclusa en España (*Anuario Estadístico del Ministerio del Interior. Año 2014*) los centros de implantación en los próximos 3 años con el sistema son los que recoge el siguiente plan.

Tabla 5-1. Plan de implantación

Áreas y/o centros de implantación	2015		2016		2017	
	S1	S2	S1	S2	S1	S2
CT7. Castilla-La Mancha y Extremadura						
CT5. Comunidad Valenciana, Murcia e Islas Baleares						
CT10. Sevilla, Cadiz y Huelva						
CT8. Asturias, Cantabria y País Vasco						
CT4. Comunidad de Madrid (*)						

(*) Los centros de Madrid I. Mujeres. Alcalá de Henares; Madrid II. Alcalá de Henares; Madrid III. Valdemoro y Madrid IV. Navalcarnero

5.1 Riesgos

Los riesgos identificados al inicio del proyecto, con sus medidas y planes de respuesta que además, han sido tenidos en cuenta para cuantificar los costes son los siguientes.

Tabla 5-2. Riesgos iniciales identificados en el proyecto

Ident	Descripción	Impacto	Probabilidad	Criticidad	Estrategia	Acciones - Planes
PDJ-RIE.001	Cambios organizativos en las instituciones. Rotación de personal clave	Bajo	Alta	Medio	Aceptar	Preparar planes específicos para velar por la estabilización presupuestaria y cumplimiento de los objetivos establecidos inicialmente.
PDJ-RIE.002	Continuas solicitudes de cambio por parte de las instituciones o agentes externos (cambios de prioridad, inclusión de nuevos trabajos, etc.)	Medio	Alto	Alta	Mitigar	Estimarlas correctamente para adaptarlas al Plan de Proyecto y solicitar su priorización por parte de la Institución
PDJ-RIE.003	Deficiencias de coordinación entre los diferentes sistemas de información. Complejidad integración de las fuentes	Alto	Bajo	Medio	Mitigar	Gestión de los recursos para elaborar un plan detallado de las diversas fuentes de información, volumetría e interoperabilidad entre sistemas.
PDJ-RIE.004	Falta de definición en el alcance y requisitos de algunas de las peticiones de los usuarios de la AP	Bajo	Baja	Bajo	Evitar	Reuniones con el usuario para delimitar el alcance y requisitos de esas peticiones
PDJ-RIE.005	Incertidumbre en las partidas presupuestarias al inicio del ejercicio dentro del programa para cubrir otras necesidades de la Administración	Medio	Baja	Medio	Aceptar	Seguimiento y control del presupuesto y revisión de los acuerdos de servicio acordados con la Administración
PDJ-RIE.006	Estimación errónea de la planificación para la implantación y el despliegue	Alta	Baja	Alta	Evitar	Seguimiento de la línea base del proyecto y elaborar las medidas correctoras necesarias para el cumplimiento de los compromisos establecidos con la Administración

6 Business Case

Manejando los **Presupuestos Generales del Estado (PGE) de 2015**, tratamos de redirigir de forma inteligente, todas aquellas partidas presupuestarias destinadas a la reinserción de los presos. El fuerte valor social de este proyecto, hace que nos centremos en el análisis desde el punto de vista de la administración, analizando su problemática como un proyecto de servicio público, contemplando nuestra colaboración público-privada como un estudio de pre-viabilidad.

Con una inversión inicial baja y con los costes fijos que consideramos que tiene la implantación de **Predictive Data Justice**, se obtiene una rentabilidad elevada con un amplio alcance económico e impacto social en los próximos 4 años.

El **porcentaje medio de reincidencia delictiva actual, ronda el 55%**, aunque depende de la tipología del delito. Diversos estudios e informes, como el presentado por el **Colegio Oficial de Psicólogos de Madrid**, dictaminan que siguiendo fehacientemente el criterio de los informes psicológicos, en algunos tipos de delitos se podría reducir hasta en un 44%. Del mismo modo en Chile una política acertada de reinserción social podría reducir hasta un 32%.

Teniendo en cuenta que nuestro modelo analítico **ha clasificado correctamente un 75% de los reincidentes** mediante la **curva ROC** del análisis IFRIP (Ver **Análisis del IFRIP** y **Medida de clasificación del modelo**) **las reducciones medias previstas de la tasa de reincidencia en los próximos 4 años-** teniendo en cuenta que se implanta en parte del territorio nacional a partir de 2016- son **3%, 7,5% y 16%** respectivamente.

Tabla 6-1. Reducción Tasa de reincidencia

	Tasa de reincidencia	
	Reducción	Tasa final
2015 - Año 1		55%
2016 - Año 2	3%	52%
2017 - Año 3	7,50%	44,50%
2018 - Año 4	16%	28,50%

6.1 Estructura de costes

La inversión que debe realizar la Administración Pública en la implantación de **Predictive Data Justice** durante los 4 próximos años es la siguiente

Tabla 6-2. Costes de implantación Predictive Data Justice

Costes de Implantación	2015	2016	2017	2018	Totales
Equipamiento Hw	116.100	35.100	31.300		182.500
Redes y Comunicaciones	58.400	40.000	22.000		120.400
Consultoría	377.320	350.840	256.760		984.920
Soporte y Mantenimiento		62.920	55.600	193.240	311.760
Formación	7.100	7.900	16.150		31.150
Otros (*)	4.150	4.150	4.650	2.150	15.100
Total por periodo	563.070	500.910	386.460	195.390	1.645.830

(*) Alojamiento y Transporte

La inversión asciende a **1.645.830 €** en la implantación del producto, por lo que se puede considerar que es relativamente **baja** respecto al beneficio que se puede llegar a obtener, tanto a nivel **económico** como a nivel **social**

Por otra parte, en el análisis para la Administración se ha tenido en cuenta dentro de las partidas presupuestarias, los **costes para programas y trabajos de reinserción**. Estos costes son variables y varían en función de la reducción de reincidencia delictiva que proporciona la implantación de **Predictive Data Justice**.

Tabla 6-3. Costes de la Administración en programas de reinserción

Costes Acciones Reinserción	2015	2016	2017	2018	Totales
Conciertos asistencia sanitaria	9.697.947	9.382.763	8.970.601	8.146.275	36.197.585
Trabajos penitenciarios. Formación empleo	6.726.951	6.508.325	6.222.430	5.650.639	25.108.345
UNED	192.885	186.616	178.419	162.023	719.943
Total por periodo	16.617.783	16.077.705	15.371.449	13.958.937	62.025.873

6.2 Rentabilidad financiera

Tomando como punto de partida los **Presupuestos Generales del Estado (PGE) de 2015** destinado a **Instituciones Penitenciaria** que ascienden a **1.124.663.000,95 €**, y con los objetivos de reducción de las tasas de reincidencia anteriormente comentados, la rentabilidad en un plazo de 4 años gracias a la implantación de **Predictive Data Justice** es la que refleja los siguientes cuadros.

Tabla 6-4. Indicadores de Rentabilidad

Indicador	Valor
ROI	54,28%
Pay-back	24 meses
VAN *	23.152.628 € * k=10% de tasa de interés
TIR	52,38%

Tabla 6-5. Resumen Flujo Neto

	Ingresos	Costes e inversiones	Flujo Neto
2015 - Año 1		17.393.813	-17.393.813
2016 - Año 2	20.103.351	16.694.495	3.408.857
2017 - Año 3	26.288.998	15.902.909	10.386.089
2018 - Año 4	52.577.995	14.160.097	38.417.898

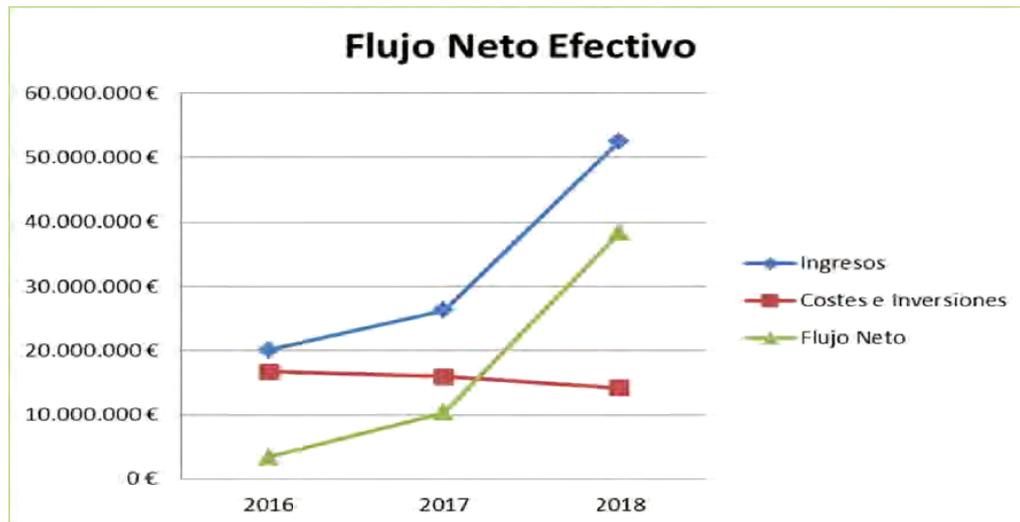


Figura 6-1. Flujo Neto Efectivo

7 Conclusiones

El objetivo principal de la solución **Predictive Data Justice** consiste en proveer de un sistema basado en inteligencia de negocio a la **Secretaría General de Instituciones Penitenciarias (SGIIPP)** y entidades colaboradoras, con el objetivo de reducir el porcentaje de reclusos que cometen reincidencia delictiva, a través de un modelo analítico-predictivo que establece una serie de mejores recomendaciones basadas en posibles tratamientos (Next Best Action) personalizables y adecuadas para cada tipo de recluso para lograr su reinserción.

España tiene una tasa de reclusos de 153/100.000 habitantes, más del doble de la Unión Europea y una tasa de reincidencia delictiva superior al 50% por lo que el gasto incurrido por la **Administración General del estado (AGE)** es bastante elevado

Predictive Data Justice proporciona una mejor adecuación de la partida presupuestaria destinadas a estas materias. Es un producto cuya implantación es económicamente viable y rentable ya que su inversión es relativamente baja con respecto al retorno que se puede llegar a obtener, **destacando el fuerte valor y dimensión social** con la mejor solución tecnológica actualmente existente, obteniendo unos porcentajes de reducción en 4 años cercanos al 16%, lo que supone un retorno de la inversión cercano al 55%, y un ahorro a las arcas públicas de más de 50 millones de euros.

8 Mejoras

En el diseño del producto **Predictive Data Justice** se han detectado nuevos puntos a desarrollar en el futuro lo que le convertirá, sin duda, en una referencia dentro del sector. Se pueden resumir en los siguientes puntos:

- ✓ **Sistema Cognitivo.**
- ✓ **Plataforma multilinguaje.**
- ✓ **Inclusión en el modelo predictivo de fuentes multimedia (video y audio).**

9 Glosario de términos

AGE	Administración General del Estado
CENDOJ	Centro de Documentación Judicial
EJE	Expediente Judicial Electrónico
FyCSE	Fuerzas y Cuerpos de Seguridad del Estado
HCS	Historial Criminológico y Social
IFRIP	Inventario de Factores de Riesgo e Intervención en Prisiones
MI	Ministerio del Interior
MJ	Ministerio de Justicia
ONG	Organizaciones No Gubernamentales
PGE	Presupuestos Generales del Estado
PMP	Project Management Professional
SGIIPP	Secretaría General de Instituciones Penitenciarias
SISPE	Sistema Informático Social Penitenciario
