

---

# PROYECTO ROLE MINING

---

Actividad 1 Elaboración de una propuesta de proyecto de IA con enfoque empresarial

Presentada por: Javier Llorente Leyton

Asignatura: Investigación y Gestión de Proyectos en Inteligencia Artificial

Profesor: Joaquín Gaspar Medina Arco

Fecha de entrega: 12/12/2025

## TABLA DE CONTENIDOS:

<b>HISTORIAL DE REVISIONES.....</b>	<b>2</b>
<b>1 INTRODUCCIÓN.....</b>	<b>2</b>
1.1 OBJETIVO.....	2
<b>2 GLOSARIO DE TERMINOS.....</b>	<b>2</b>
<b>3 SITUACIÓN ACTUAL .....</b>	<b>3</b>
3.1 ASIGNACIÓN DE ROLES .....	3
3.2 PROBLEMA ENCONTRADO.....	3
<b>4 TRABAJOS RELACIONADOS.....</b>	<b>5</b>
<b>5 PROYECTO DE ROLE MINING .....</b>	<b>6</b>
5.1 VALIDACIÓN DEL PROYECTO .....	7
<b>6 PLAN DE IMPLEMENTACIÓN .....</b>	<b>8</b>
6.1 EQUIPO DEL PROYECTO .....	8
6.2 FASES DEL PROYECTO .....	9
6.3 COSTE DEL PROYECTO.....	12
<b>7 CONCLUSIONES.....</b>	<b>13</b>
7.1 MEJORA CONTINUA .....	13
<b>8 BIBLIOGRAFIA Y USO DE LA IA .....</b>	<b>13</b>

## HISTORIAL DE REVISIONES

versión	Date	Autor	Descripción
1.0	12-DIC-2025	Javier Llorente	Versión Inicial

## 1 INTRODUCCIÓN

Este documento contiene la justificación y la implementación del proyecto de Role Mining y despliegue del piloto.

### 1.1 OBJETIVO

El objetivo del documento es dar una explicación de cómo se va a implementar el proyecto, queda fuera del alcance la fase de mejora continua.

## 2 GLOSARIO DE TERMINOS

Glosario de términos usados en el documento.

**Role Mining:** “Role Mining is a process in Identity and Access Management (IAM) that helps define the various roles held by users in an organization through analyzing patterns of access permissions. This automated system isolates and groups together individual users that have similar responsibilities and access needs. Role Mining lessens the complexity of managing the numerous roles and permissions across a company, simplifying the access management challenges” ([IBM](#), s. f.).

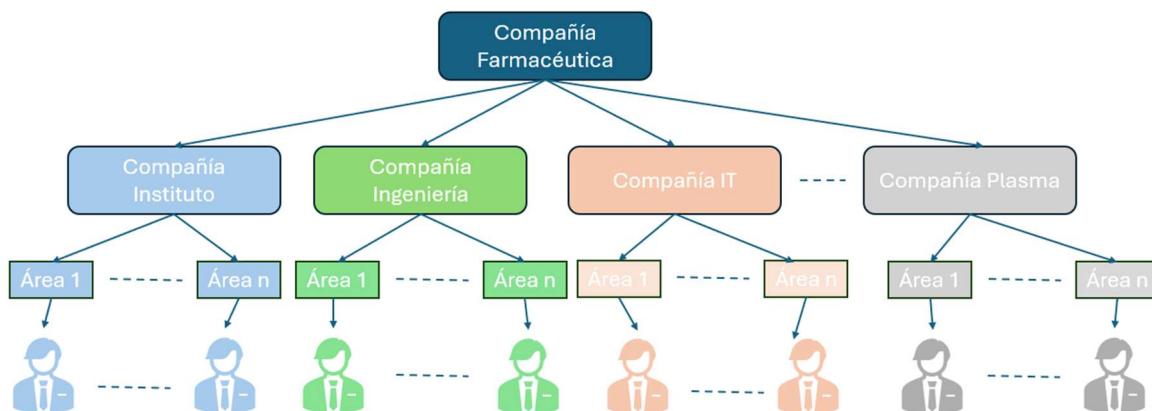
**IRMAOC:** Interpretable Role Mining Algorithm Based on Overlapping Clustering.

**ABAC:** Attribute-Based Access Control.

## 3 SITUACIÓN ACTUAL

### 3.1 ASIGNACIÓN DE ROLES

La empresa es del sector farmacéutico con +20.000 empleados, se divide en distintas compañías pertenecientes a ella y estas se subdividen en áreas (ver [Imagen 1](#)). Esto lleva a tener un número muy elevado de aplicaciones corporativas y roles asociado a las mismas. Lo que con lleva a que sea un proceso muy complejo la asignación de roles cuando un empleado entra en la empresa y el control tanto a nivel de seguridad como a nivel regulatorio.



[Imagen 1 : Organización de la empresa](#)

### 3.2 PROBLEMA ENCONTRADO

Una vez visto este problema se ha decidido implementar un proyecto de [Role Mining](#), el cual nos ayudará a facilitar la asignación de roles, el control de seguridad y la normativa. Para ello se implementará una IA mediante aprendizaje supervisado que clasificara los roles y se adaptará a los cambios organizativos que se produzcan.

Los beneficios que podemos cuantificar de la implementación de este proyecto serían los siguientes: (simulación con [OpenAI 2025](#))

1. Reducción del tiempo de certificación de los accesos.
  - a. Asignación roles manual  $20.000 \text{ empleados} \times 30 \text{ minutos} = 10.000 \text{ horas}$ , reducción de 70-90% horas.

2. Reducción del tiempo en realizar recertificaciones y SOX (Auditorias).
3. Menos accesos indebidos, esto evitaría incidentes, sanciones regulatorias (GxP, SOX).
4. Productividad nuevos empleados, un Onboarding manual puede generar de 1 a 7 días perdidos.
  - a. Estimamos unas 2000 altas año \* 1 día \* 8h \* 50 euro/día = 800.000 €

De todo esto tenemos la siguiente tabla:

Concepto	Ahorro estimado (€)
<b>Asignación roles</b>	200.000
<b>Reducción de recertificaciones y SOX</b>	180.000
<b>Menos accesos indebidos</b>	50.000
<b>Productividad nuevos empleados</b>	800.000
<b>Siendo conservador el total ahorro</b>	1.230.000

Tabla 1: Ahorro Estimado

Si observamos la [tabla 1](#) donde tenemos el ahorro estimado podemos calcular el tiempo de retorno de la inversión y el ROI para un coste de proyecto de 500.000 €.

**Retorno de la inversión:**  $500.000 / 1.230.000 = 0,40$  años  $\rightarrow 4,87$  meses

**ROI anual:** 146 %

## 4 TRABAJOS RELACIONADOS

Algunos de los papers relacionados con Role Mining, fueron identificados utilizando [OpenAI 2025](#).

En estos trabajos (ver [Bibliografía](#)) abordan diferentes enfoques para implementación de Role Mining y modelos de control de acceso. Entre los 5 trabajos analizados, tres de ellos presentan implementaciones con los métodos [IRMAOC](#), [HTPA](#) los cuales realizan el proceso de agrupamiento sin apoyarse en técnicas de inteligencia artificial, estos trabajos ofrecen algoritmos tradicionales bien estructurados que permiten obtener roles a parte ir de permisos y asignaciones de usuarios, basándose principalmente y métodos estadísticos, heurísticos o de optimización.

Por otra parte, los otros dos trabajos se centran en la aplicación del modelo ABAC, aunque ABAC no corresponde directamente a un proceso de Role Mining, resulta relevante debido a que sus mecanismos de control basados en atributos pueden complementar lo.

Estos trabajos aportan algoritmos funcionales y robustos para el descubrimiento y definición de roles. Lo que aporta la Inteligencia artificial es una reducción del ruido, obtención de roles seguros y coherentes, mejoras en comprensión de los resultados al poder dar respuestas más naturales y una mayor adaptabilidad a los cambios.

## 5 PROYECTO DE ROLE MINING

En este proyecto de role Mining plantea una solución basada en inteligencia artificial que, mediante técnicas de aprendizaje supervisado, permitirá clasificar los roles existentes y se adaptarse de forma dinámica a los cambios organizativos de la empresa.

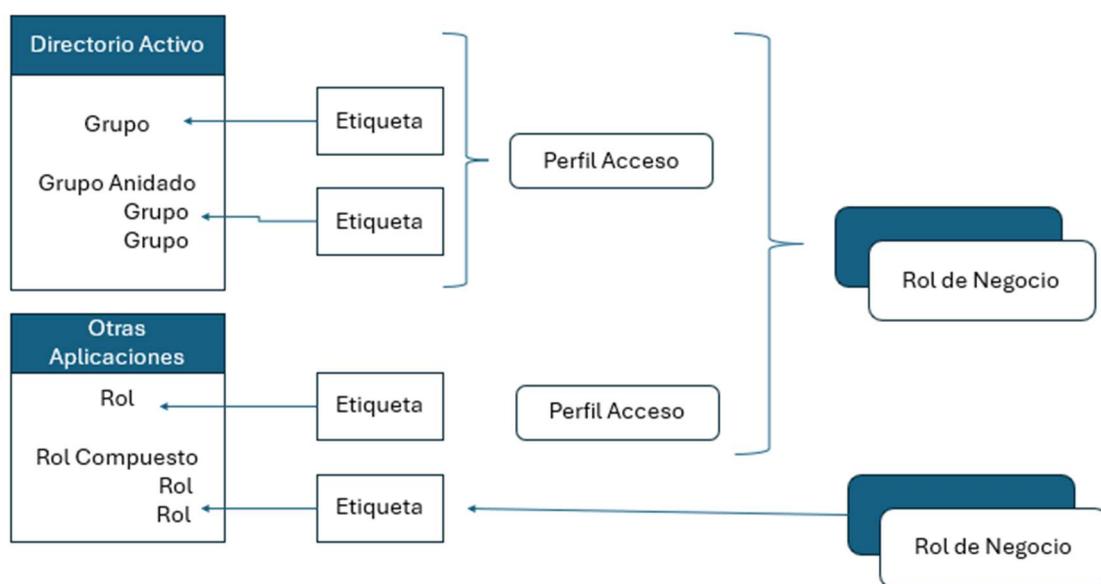


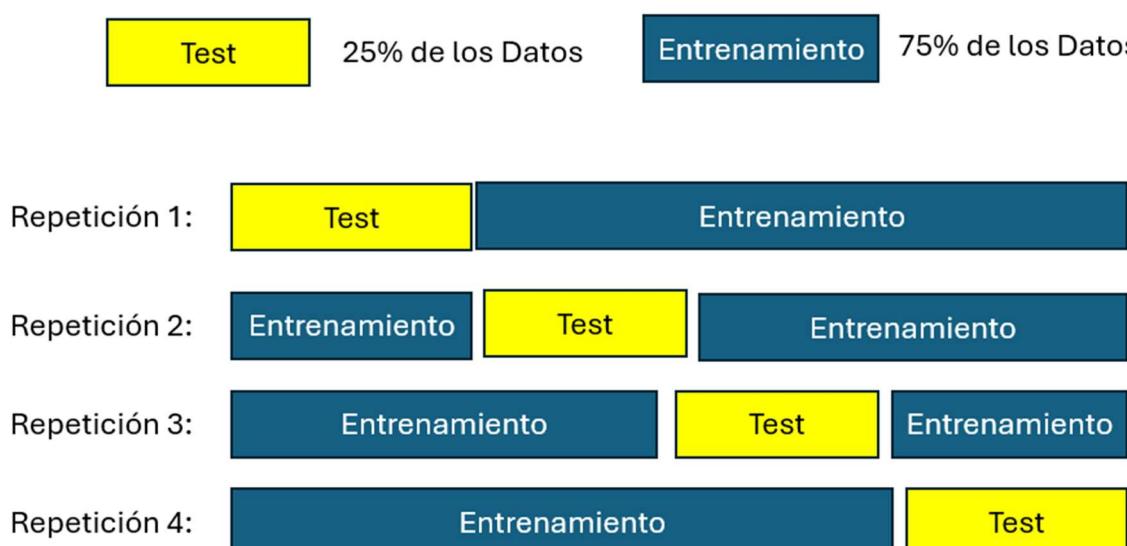
Imagen 2: Role Mining

El objetivo principal del proyecto es simplificar la asignación de roles, eliminar redundancias, proporcionar un mayor control en la seguridad y en el cumplimiento de la normativa del sector farmacéutico.

Para el entrenamiento del modelo de IA se utilizará un conjunto de datos recopilados durante las fases del proyecto comprensión del negocio y comprensión de los datos. En estas fases se tendrán reuniones con los stakeholders para identificar y documentar los roles de negocios conocidos. Esta información servirá como entrada al sistema a la IA, los distintos roles de Active Directory (LDAP) y de las distintas aplicaciones corporativas. Como resultado de la IA, generará agrupaciones coherentes de roles de negocio (ver [Imagen 2](#)) que serán los que se asignarán a los empleados.

## 5.1 VALIDACIÓN DEL PROYECTO

La validación del proyecto se realizará mediante un subconjunto de test que se sacará del conjunto de entrenamiento de la IA. Para garantizar la robustez del modelo se aplicará el método de validación cruzada (ver [Imagen 3](#)), lo que permitirá comprobar que la IA clasifica correctamente los roles y que no existe sobreajuste, es decir, el modelo no se limita a responder correctamente al conjunto de entrenamiento.



[Imagen 3: Validación cruzada.](#)

Se considerará que el piloto está listo para el despliegue cuando cumpla los siguientes criterios:

1. Los roles sugeridos se han aceptables por negocio en un 90%.
2. Los permisos inseguros detectados hayan sido corregidos.

## 6 PLAN DE IMPLEMENTACIÓN

La metodología que aplicaremos para este proyecto será Agile, utilizando concretamente el CRISP-DM. Esta metodología da la flexibilidad necesaria para un proyecto de Inteligencia artificial, porque permite iterar entre todas las fases del proyecto hasta llegar a la fase final del despliegue del piloto. Una vez desplegado el piloto entraremos en una fase de mejora continua.

### 6.1 EQUIPO DEL PROYECTO

**Project Manager:** Responsable de realizar el seguimiento del proyecto y de su metodología (Scrum Máster).

**Responsable de negocio:** Actuará como enlace principal con los stakeholders, asegurando que las necesidades del negocio queden correctamente representadas.

**Especialista roles:** Persona encargada de conocer y/o recopilar los roles existentes en las distintas aplicaciones, con los responsables de estas.

**Especialista en Datos:** Será el encargado de analizar y preparar los datos necesarios para la creación del modelo en el entrenamiento de la IA.

**Especialista en IA:** Responsable de ejecutar la fase de entrenamiento de la IA y la de test junto al especialista en datos.

**Responsable Calidad:** Velará por el cumplimiento de los requisitos GxP y SOX.

**Responsable Infraestructura:** Responsable de crear la infraestructura necesaria para poder hacer el despliegue del proyecto y despliegue de este.

## 6.2 FASES DEL PROYECTO

Todas las fases se podrán repetir para adaptar el proyecto a los resultados obtenidos en cada fase posterior hasta llegar a la fase final en que se desplegará el piloto.

El proyecto tendrá una duración de 6 meses por lo que las iteraciones tendrán que estar dentro de los 6 meses descontado el tiempo necesario de despliegue. Si se llega al tiempo máximo se tendrá que validar si podemos desplegar el piloto o tenemos que replanificar el proyecto (Ampliando duración, presupuesto, etc.)

### Compresión del Proyecto

Definir los objetivos del proyectos y criterios de éxito:

Tareas	Entregables	Responsables	Duración
<ul style="list-style-type: none"> <li>• Reuniones con stakeholders.</li> <li>• Identificación Objetivos</li> <li>• Definición de alcance y riesgo.</li> </ul>	<ul style="list-style-type: none"> <li>• Documentos de objetivos.</li> <li>• Plan de riesgos Inicial</li> </ul>	<ul style="list-style-type: none"> <li>• Project Manager.</li> <li>• Responsable de negocio</li> </ul>	1 o 2 semanas

Tabla 2: Fase de comprensión del proyecto

### Comprensión de los datos

Recopilar y explorar permisos:

Tareas	Entregables	Responsables	Duración
<ul style="list-style-type: none"> <li>• Obtención de los roles.</li> <li>• Análisis de los datos.</li> </ul>	<ul style="list-style-type: none"> <li>• Informe de los roles obtenidos.</li> <li>• Dataset Inicial.</li> </ul>	<ul style="list-style-type: none"> <li>• Especialista roles</li> <li>• Especialista en Datos</li> </ul>	1 o 2 semanas

Tabla 3: Fase de comprensión de los datos

Preparación de los datos

Limpiar y transformar datos para el análisis:

Tareas	Entregables	Responsables	Duración
<ul style="list-style-type: none"> <li>• Normalización de permisos.</li> <li>• Eliminación de datos redundantes y ausentes.</li> <li>• Detección de anomalías.</li> </ul>	<ul style="list-style-type: none"> <li>• Dataset final.</li> <li>• Informe de estado de los datos</li> </ul>	<ul style="list-style-type: none"> <li>• Especialista roles.</li> <li>• Especialista en Datos.</li> </ul>	1 o 2 semanas

Tabla 4: Fase de preparación de datos

Modelado

Aplicar el aprendizaje supervisado para descubrir los roles de negocio:

Tareas	Entregables	Responsables	Duración
<ul style="list-style-type: none"> <li>• Clasificación de roles.</li> <li>• Validación del modelo.</li> </ul>	<ul style="list-style-type: none"> <li>• Modelo de roles propuestos.</li> <li>• Informe de la validación.</li> </ul>	<ul style="list-style-type: none"> <li>• Especialista en IA.</li> <li>• Especialista en Datos.</li> </ul>	1 o 2 semanas

Tabla 5: Fase de Modelado

Evaluación

Validar roles propuestos con roles reales.

Tareas	Entregables	Responsables	Duración
<ul style="list-style-type: none"> <li>• Validación funcional.</li> <li>• Validación calidad</li> </ul>	<ul style="list-style-type: none"> <li>• Revisión con expertos funcionales.</li> <li>• Documento aceptación calidad.</li> </ul>	<ul style="list-style-type: none"> <li>• Especialista en IA</li> <li>• Especialista en Datos</li> <li>• Responsable Calidad</li> </ul>	1 o 2 semanas

Tabla 6: Fase de Evaluación

Despliegue

Implementar el proyecto y monitorizarlo.

Tareas	Entregables	Responsables	Duración
<ul style="list-style-type: none"> <li>• Despliegue del piloto</li> </ul>	<ul style="list-style-type: none"> <li>• Informe post- implementación.</li> <li>• Documento de despliegue en producción</li> </ul>	<ul style="list-style-type: none"> <li>• Project Manager.</li> <li>• Responsable de negocio</li> <li>• Responsable Infraestructura</li> <li>• Especialista en Datos</li> <li>• Responsable Calidad</li> </ul>	2 semanas

Tabla 7: Fase de despliegue

### 6.3 COSTE DEL PROYECTO

El coste del Proyecto simulado con [OpenIA 2025](#).

El Proyecto tendrá una duración de 6 meses 120 días laborables.

Rol	Tarifa
<b>Project Manager</b>	800 €/día
<b>Responsable Negocio</b>	500 €/día
<b>Especialista en roles</b>	500 €/día
<b>Especialista en datos</b>	600 €/día
<b>Especialista en IA</b>	800 €/día
<b>Responsable calidad</b>	400 €/día
<b>Responsable Infraestructura</b>	400 €/día

Tabla 8: Tabla de tarifas.

Rol	Días asignados	Costes
<b>Project Manager</b>	120	96.000 €
<b>Responsable Negocio</b>	120	60.000 €
<b>Especialista en roles</b>	120	60.000 €
<b>Especialista en datos</b>	120	72.000 €
<b>Especialista en IA</b>	90	72.000 €
<b>Responsable calidad</b>	50	20.000 €
<b>Responsable Infraestructura</b>	50	20.000 €

Tabla 9: Coste del proyecto

Con los datos de las tablas [8](#) y [9](#) tenemos:

Costes adicionales: Licencias 100.000 €

Coste total del Proyecto 500.000 €

## 7 CONCLUSIONES

### 7.1 MEJORA CONTINUA

Una vez se despliegue el proyecto entraremos en la fase de mejora continua donde tendremos los siguientes objetivos:

- Monitorización del funcionamiento del piloto desplegado para asegurar su rendimiento y detectar posibles incidencias.
- Ajuste y actualización de roles a cambios organizativos o nuevas necesidades.
- Integración con la aplicación de ticketing, en nuestro caso Service Now, para automatizar y optimizar la gestión de solicitudes relacionadas con roles y permisos.

## 8 BIBLIOGRAFIA Y USO DE LA IA

- IBM. (s. f.). *What is role Mining?* <https://www.entitle.io/resources/glossary/role-mining>
- Yang, Y., Li, J., Zhang, T., Chen, L., Huang, G., & Lv, Z. (2025). *IRMAOC: an interpretable role Mining algorithm based on overlapping clustering*. *Cybersecurity*, 8, 54. <https://doi.org/10.1186/s42400-024-00348-z>
- Bui, T., Shabram, E., & Matricia, A. (2025). *An approach for handling missing attribute values in attribute-based access control policy mining*. arXiv. <https://arxiv.org/abs/2505.01873>
- Didinova, I., & Macek, K. (2024). *Identification and Attribution of Access Roles Using Hierarchical Team Permission Analysis*. En *Proceedings of the 26th International Conference on Enterprise Information Systems (ICEIS 2024)*. SCITEPRESS. <https://doi.org/10.5220/0012675300003690>
- Crampton, J., Eiben, E., Gutin, G., Karapetyan, D., & Majumdar, D. (2024). *Bi-objective Optimization in Role Mining*. arXiv. <https://arxiv.org/abs/2403.16757>

- — *ABAC policy mining method based on hierarchical clustering and relationship extraction.* (2024). *Computers & Security*, 139, 103717. <https://doi.org/10.1016/j.cose.2024.103717>
- Trabajos relacionados. OpenAI. (2025). ChatGPT (GPT-5.1) [Modelo de lenguaje grande]. <https://chatgpt.com/share/6939574b-3a80-8013-8ae5-c90d498252b5>
- Simulación ROI. OpenAI. (2025). ChatGPT (GPT-5.1) [Modelo de lenguaje grande]. <https://chatgpt.com/share/693947a2-f7fc-8013-880f-9f0dc9075b44>
- Simulación Costes. OpenAI. (2025). ChatGPT (GPT-5.1) [Modelo de lenguaje grande]. <https://chatgpt.com/share/69394938-8090-8013-8709-8d1f315d371f>