# Sistema de Detección de Valores Atípicos y Evaluación de Riesgos en la Gestión de Acceso a SaaS

## Temática General

## El proyecto se centra en el desarrollo de un sistema para la detección de valores atípicos en un conjunto de datos. Este conjunto de datos representa la jerarquía de empleados de una empresa y sus accesos a distintas aplicaciones de software. El objetivo es determinar cuál de las técnicas y bibliotecas de código abierto disponibles tiene un mejor rendimiento a la hora de localizar grupos de accesos atípicos.

## Objetivos del Proyecto

### Objetivo Principal

Construir un sistema capaz de identificar "outliers", valores atípicos, o características que se desvíen significativamente del resto. La identificación de estos valores atípicos se realizará en diferentes dimensiones, tales como:

* Líneas de reporte de empleados y jerarquía de la empresa.
* Grupos de usuarios creados por los gerentes y administradores.
* Accesos o cuentas SaaS pertenecientes a los empleados, con su respectivo nivel de permiso dentro del software.
* Tamaño de la empresa.

## Requisitos del Proyecto

### Stack tecnológico

El desarrollo del proyecto requerirá la utilización de las siguientes tecnologías:

* *Backend*: Construcción de una WebAPI utilizando .NET Core para manejar la orquestación, lógica del negocio y la interacción con base de datos.
* *Procesamiento de datos*: Desarrollo de un motor para la identificación de los outliers usando tecnologías basadas en python como Scikit-learn, SciPy, HDBSCAN, TensorFlow, PyTorch u otros.

### Datos y Base de Datos

Se dispone de una base de datos anonimizada que contiene información de diversas dimensiones, mencionadas previamente. La gestión y el análisis de esta base de datos es crucial para el desarrollo del sistema.

### Estrategias y Métodos

El proyecto debe explorar y evaluar distintas estrategias para la identificación de valores atípicos. Esto incluye, pero no se limita a:

* *Inteligencia Artificial & Clustering*: Agrupamiento de datos basado en similitudes y diferencias para identificar patrones y anomalías. Implementación de algoritmos de aprendizaje automático y técnicas avanzadas de IA para la detección de outliers.
* *Base de Conocimiento*: Utilización de un sistema basado en reglas predefinidas que puedan detectar anomalías.
* *Heurística Manual*: Aplicación de métodos y reglas heurísticas para identificar manualmente valores atípicos.

Es importante realizar pruebas comparativas entre al menos dos de estas estrategias para evaluar su eficacia en la identificación de outliers y la detección de riesgos de seguridad asociados.