

# AuditGov: Analítica para la Detección de Riesgo de Fraude en Entidades Públicas

## 1. Introducción

AuditGov, fundada en 1989, es una de las firmas de auditoría externa gubernamental más influyentes del país. Desde sus oficinas centrales en Nueva Delhi, la compañía supervisa la integridad financiera y operativa de 777 entidades públicas distribuidas en sectores estratégicos como Agricultura, Infraestructura, Pesca, Ciencia y Tecnología, Turismo, Salud Pública, entre otros. Su mandato es claro: **detectar señales tempranas de fraude, corrupción o errores materiales** en la gestión de recursos públicos.

A pesar de su rol estratégico, AuditGov enfrenta un reto creciente: el ecosistema gubernamental ha aumentado enormemente su complejidad en los últimos diez años. Los avances en digitalización han generado grandes volúmenes de datos, pero también han hecho más sofisticadas las prácticas fraudulentas. Las técnicas manuales que antes eran suficientes ahora resultan insuficientes para detectar patrones de riesgo ocultos entre cientos de variables y miles de registros administrativos.

Históricamente, la firma ha realizado auditorías basadas en:

- La **experiencia acumulada** de sus auditores senior.
- Señales cualitativas identificadas durante entrevistas.
- Revisión manual de reportes financieros.
- Registros de inconsistencias reportadas por sectores locales.

Aunque este enfoque permitió resultados aceptables durante décadas, hoy es evidente que la organización está rezagada respecto a estándares globales adoptados por firmas como Deloitte, EY o PwC, que ya integran sistemas de analítica predictiva y machine learning para priorizar riesgos.

En 2024, un informe independiente del Ministerio de Finanzas reveló que **el 38% de los hallazgos significativos del año anterior se produjeron en entidades que no habían sido clasificadas como de alto riesgo en la planeación**. Este hallazgo despertó preocupación en la junta directiva de AuditGov:

¿Estaba la firma evaluando correctamente el riesgo?

¿Existía sesgo o inconsistencia en la priorización manual?

¿Era tiempo de adoptar herramientas basadas en machine learning?

La firma decidió emprender un proyecto piloto aprovechando un información histórica de 777 entidades ya clasificadas como **Fraud** o **No-Fraud**, y variables de riesgo calculadas en auditorías previas. El objetivo era claro: **construir un modelo predictivo que permitiera priorizar**

**esfuerzos, disminuir la subjetividad e incrementar la eficiencia del proceso de planeación anual.**

---

## 2. Situación Actual

En enero de 2025, AuditGov inició la planeación de auditorías anuales para el siguiente ciclo fiscal. El equipo humano estaba compuesto por:

- **150 auditores internos**, distribuidos por regiones.
- **12 especialistas financieros senior**, encargados de validar criterios de riesgo.
- **1 departamento de TI** pequeño, enfocado en soporte, con poca experiencia en ciencia de datos.
- **Presupuesto limitado** que obliga a priorizar solo 120 visitas presenciales al año.

Cada año, los equipos deben decidir cuáles entidades representan la **mayor probabilidad de fraude o mala gestión**. Sin embargo, en años recientes han surgido varios problemas:

### 2.1. Sobrecarga de datos

Las entidades entregan:

- Estados financieros completos.
- Registros administrativos.
- Informes de operaciones.
- Datos estructurados y no estructurados.

En total, más de **45.000 páginas revisadas manualmente** cada año, lo que hace imposible que cada auditor profundice en todos los indicadores relevantes.

### 2.2. Inconsistencias en la aplicación de criterios

Un análisis interno reveló:

- En la región norte, el 70% de entidades agrícolas se clasificaban como alto riesgo.
- En la región sur, solo el 20% se consideraban riesgosas, pese a tener indicadores similares.

Los auditores argumentaron diferencias contextuales, pero la dirección sospecha que puede existir **subjetividad, sesgo o exceso de dependencia en la intuición**.

### 2.3. Evaluación manual fragmentada

La construcción del "Risk Audit Score" es un proceso que combina:

- Factores financieros (liquidez, gasto, variaciones).
- Factores administrativos (historial de irregularidades).
- Factores sectoriales (regulación, carga operativa, tamaño).
- Factores históricos (hallazgos previos en ese sector).

Cada factor era evaluado en una hoja Excel independiente por diferentes equipos. Esto generaba:

- **Criterios duplicados o inconsistentes.**
- Pérdida de trazabilidad.
- Falta de un repositorio unificado.
- Altísima dependencia de conocimiento tácito.

### 2.4. Variabilidad de los auditores

Los auditores más experimentados tendían a:

- Penalizar más sectores históricamente conflictivos (Agricultura, Industria).
- Subestimar sectores emergentes (Tecnología, Ciencia).
- Sobreponderar o subponderar factores subjetivos como “comportamiento del director financiero”.

En auditorías posteriores, estas decisiones demostraron tener una correlación débil con los hallazgos reales.

### 2.5. Resultados inconclusos del método PSO

AuditGov desarrolló un experimento interno aplicando **Particle Swarm Optimization (PSO)** para rankear las variables de riesgo, pero:

- No existió integración con un modelo predictivo completo.
- Los auditores no entendieron cómo interpretar los pesos aprendidos.
- La herramienta quedó subutilizada.

### 2.6. Presión regulatoria

El Ministerio de Finanzas envió una circular requiriendo:

“Mayor objetividad y evidencia empírica en la priorización de auditorías de alto impacto.”

Esto puso presión adicional en AuditGov para modernizar su proceso.

## 2.7. El dilema

El Director General de AuditGov está dividido entre dos caminos:

### *Camino A – Continuar con la metodología tradicional*

- Mantiene el criterio profesional del auditor.
- Evita costos de implementación tecnológica.
- Minimiza resistencia cultural.

### *Camino B – Adoptar un modelo de predictivo*

- Permite predecir fraude con base en datos.
- Reduce subjetividad y variabilidad regional.
- Mejora la priorización y ahorra millones en recursos.
- Pero requiere:
  - infraestructura,
  - capacitación,
  - gobernanza de datos,
  - cambio organizacional.

La junta directiva está lista para tomar una decisión, pero antes deben evaluar los resultados del piloto analítico y la recomendación del equipo consultor... rol que asumirán los estudiantes.

## 3. Problema de Negocio

Pese a contar con datos y un equipo competente, AuditGov enfrenta:

### 3.1. Limitación de recursos

- Solo pueden visitar **120 entidades al año**.
- El proceso actual de priorización toma semanas.
- Existe presión política y reputacional si se audita la entidad equivocada.

### 3.2. Ausencia de una metodología objetiva

Los auditores dependen de criterios subjetivos (“intuición”, “experiencia previa”). Esto genera inconsistencias entre regiones y equipos.

### 3.3. Demoras en la planeación anual

La planeación debe realizarse en menos de 30 días, pero el análisis tradicional puede tardar hasta 45 días.

### 3.4. Necesidad de adoptar analítica avanzada

La firma quiere moverse hacia un modelo similar a Big Four, donde las decisiones son data-driven.

---

## 4. Datos Disponibles

- 33 variables numéricas y categóricas.
- Sectores: Agricultura, Irrigación, Salud Pública, Industria, Pesca, Ciencia y Tecnología, etc.
- 777 registros (uno por entidad auditada).
- Variable objetivo: **Risk\_Audit\_Class** (Fraud / No-Fraud).
- Variables de riesgo calculadas históricamente.
- Variables estimadas mediante PSO (Particle Swarm Optimization) que rankean la importancia de los factores.

Los datos provienen del estudio realizado por Houston, Peters, y Pratt (1999) sobre factores de riesgo y comportamiento del fraude en organizaciones públicas.

---

## 5. Decisión Estratégica

El comité de auditoría debe decidir si:

1. **Implementar un sistema predictivo basado, o**
2. Continuar con el método tradicional de planeación manual.

La gerencia solicita un **análisis técnico y de negocio** que permita justificar la adopción del sistema para mejorar la priorización del riesgo.

---

## 6. Preguntas del Caso

### 6.1. Diagnóstico

- ¿Qué variables parecen estar asociadas con fraude?
- ¿Cómo es la distribución de sectores y cuál podría influir más en el riesgo?

### 6.2. Modelamiento

- Entrenar al menos tres modelos (por ejemplo: **Logistic Regression, Random Forest, XGBoost**).
- Comparar su desempeño (accuracy, F1-score, ROC-AUC,...).
- Determinar cuál modelo recomendar a AuditGov.

### 6.3. Interpretabilidad

- Identificar las variables más importantes según el modelo ganador.
- Explicar cómo impactan el riesgo de fraude.
- ¿Cómo se pueden presentar estos resultados a auditores no técnicos?

### 6.4. Priorización de Auditorías

- A partir de las predicciones del modelo, seleccionar las **120 entidades de mayor riesgo**.
- Justificar la recomendación de priorización.

### 6.5. Recomendación ejecutiva

- ¿Cómo debe la empresa integrar este modelo en su proceso de planeación?
- ¿Qué riesgos éticos existen al usar ML en procesos de auditoría pública?
- ¿Qué pasos seguir para operacionalizar este sistema dentro del ciclo anual?

---

## 7. Entregables Esperados

1. **Informe ejecutivo** para la junta directiva (1-2 páginas).
2. **Notebook de ML** con el análisis exploratorio, entrenamiento y evaluación.
3. **Ranking de 120 entidades recomendadas para auditoría.**
4. **Dashboard o reporte visual** para los auditores.

## 5. Recomendaciones de política y gobernanza de datos.

---

## 8. Cierre del Caso

AuditGov se encuentra en un punto crítico.

La presión regulatoria aumenta, los riesgos son complejos y los recursos limitados.

La organización sabe que necesita **una herramienta inteligente** que permita identificar señales tempranas de fraude y optimizar la asignación de recursos.

De la calidad de tu análisis dependerá si la firma adopta el sistema inteligente como herramienta principal para la planificación de auditoría.