

---

# La analítica avanzada de datos en la Seguridad Social.

**La Seguridad Social se enfrenta en la actualidad a diversos retos de gran envergadura entre los que se encuentran el garantizar la sostenibilidad del sistema público de pensiones en el medio y largo plazo, la revisión del modelo de atención a ciudadanos, empresas y trabajadores como resultado de la aplicación de las Leyes 39 y 40 o la lucha contra el fraude en el cobro de las prestaciones.**



---

**JUAN PARDO GARCÍA**

Jefe de Área de Estadísticas y Análisis de Datos

Gerencia de Informática de la Seguridad Social

Para afrontar dichos retos es fundamental disponer de herramientas que faciliten en tiempo y forma toda la información necesaria para apoyar los procesos de toma de decisión por parte de las Entidades Gestoras y el resto de órganos directivos adscritos a la Secretaría de Estado de la Seguridad Social (SESS).

La Gerencia de Informática de la Seguridad Social (GISS), como Centro Directivo responsable de proporcionar los servicios TIC en el ámbito de la SESS, provee a través de su plataforma analítica las capacidades necesarias en materia de análisis de datos en sus distintas vertientes<sup>1</sup>.

Por un lado, la aplicación de técnicas tradicionales de análisis de datos como pueden ser el reporting básico o los cuadros de mando permite disponer de una visión retrospectiva de la organización para dar respuesta a preguntas como “qué ha pasado” o el “por qué” se ha llegado la situación actual.

Pero más allá de facilitar la toma de decisiones mediante el análisis descriptivo de los datos, la aplicación de técnicas avanzadas de aprendizaje automático o “Machine Learning” como el análisis predictivo y prescriptivo posibilita el disponer de una visión a futuro de la organización donde lo que se busca no es ya tanto apoyar la toma de decisiones, sino la optimización de los procesos de negocio y el incremento de la productividad de las distintas unidades administrativas.

# Monográfico



**Figura 1. Analítica tradicional vs. Analítica avanzada**

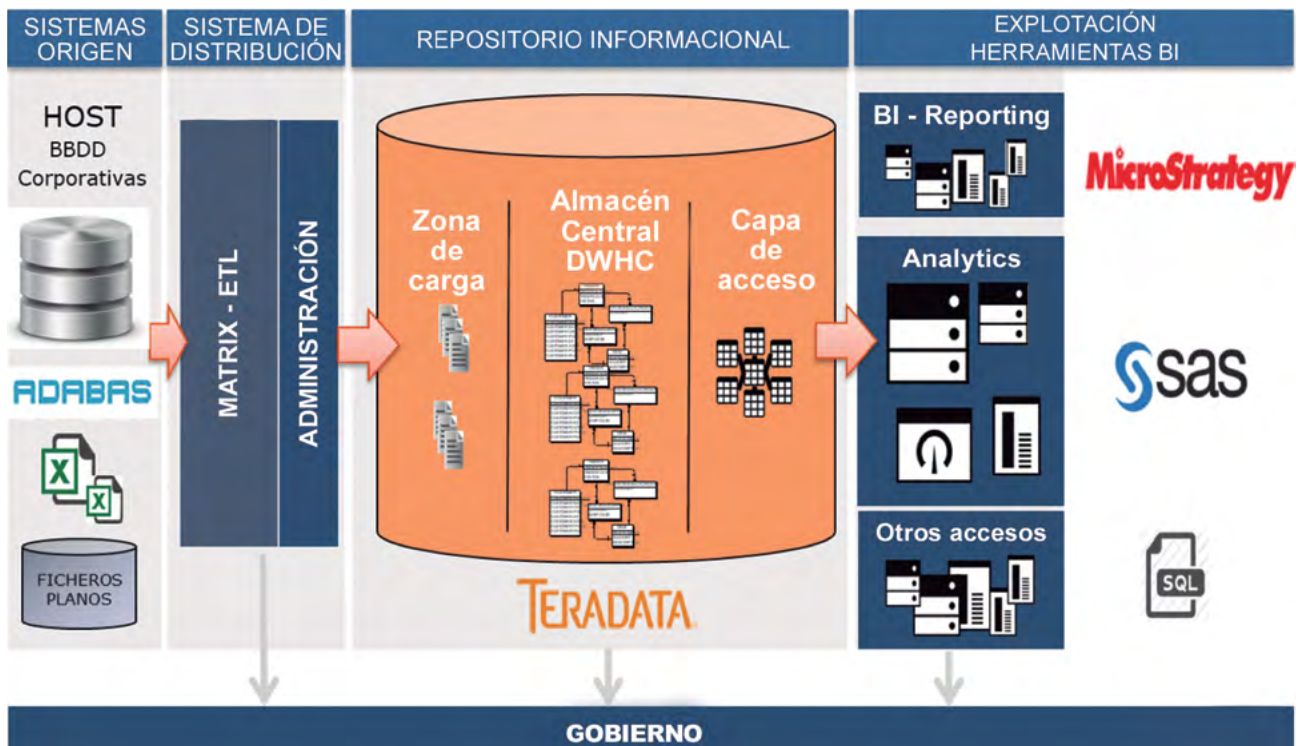
## La plataforma analítica de la Seguridad Social

El entorno analítico de la Seguridad Social conforma un repositorio informacional con más de 170.000 millones de registros y una ocupación que supera los 25 TB en el que, entre otros, se recogen datos de Afiliación relativos a las relaciones laborales existentes entre 18,5 millones de

trabajadores activos y 1,5 millones de empresas, datos de la Recaudación de cuotas a la Seguridad Social que genera unos 70 millones de registros mensuales de bases de cotización, compensaciones y bonificaciones, o datos de la nómina de las 9,5 millones de Pensiones vigentes y de los más de 4 millones de procesos de Incapacidad Temporal informados anualmente

por los Servicios Públicos de Salud de las Comunidades Autónomas.

Para gestionar dicho volumen de información, la GISS inició en el año 2014 un proceso de transformación y consolidación del entorno analítico con la unificación de las diversas soluciones departamentales que existían hasta entonces en el ámbito de las distintas Entidades Gestoras. De este modo, y siguiendo el paradigma de separación de entornos, se estableció un repositorio informacional único para toda la SESS, independiente de los sistemas transaccionales de origen, que está basado en Teradata. Éste se alimenta a través de un sistema de distribución o ETL propio denominado MATRIX y como herramientas de explotación se utilizan Microstrategy para el BI clásico y SAS Fraud Framework para la analítica más avanzada. La plataforma se completa con subsistema transversal orientado al gobierno del dato que facilita la trazabilidad del mismo desde los sistemas de origen hasta las aplicaciones de explotación<sup>2</sup>.



**Figura 2. La plataforma analítica de la Seguridad Social**

En la actualidad, la GISS se encuentra en proceso de ampliación de las capacidades del entorno analítico con la incorporación de una plataforma de Big Data-Hadoop que permita realizar el tratamiento de datos no estructurados como, por ejemplo, documentos de sentencias judiciales, convenios colectivos, logs de portales de internet, etc.

### Proyectos de analítica avanzada de datos

La existencia de un repositorio único ha permitido integrar la información proveniente de los distintos ámbitos de negocio, resolviendo el problema que suponía la existencia de silos de información aislada en los distintos sistemas departamentales. En torno a dicho repositorio único se están desarrollando diversos proyectos de análisis predictivo cuya finalidad es ayudar a optimizar el trabajo de determinadas unidades de gestión como la Inspección de Trabajo, las Unidades de Inspección Médica o las Inspecciones de Servicios, cuya carga de trabajo excede, en muchos casos, la capacidad operativa de las mismas.

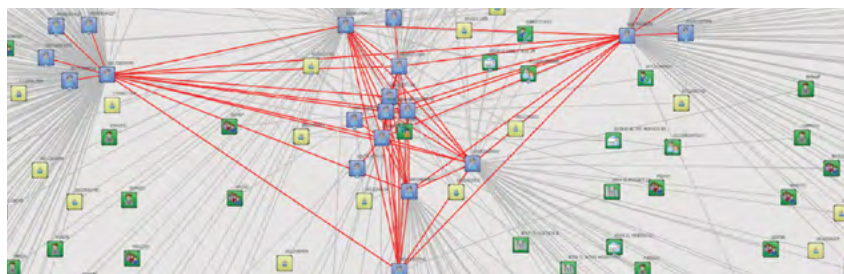


Figura 4. Visualización de redes de fraude.

Para abordar estos proyectos, la GISS está siguiendo las recomendaciones de la metodología SEMMA de SAS<sup>3</sup>. De este modo, en las primeras fases de cada proyecto se realiza una preparación de los datos, muestreando y seleccionando las variables relevantes cuyo resultado es la construcción de un tablón o Datamart Analítico. Posteriormente, se aplican sobre éste distintos modelos de regresiones, árboles de decisión, redes neuronales, etc., se elige el que ofrece mejores resultados predictivos y finalmente se ejecutan y validan los resultados sobre un escenario real.

### Analítica Predictiva en la Lucha contra el Fraude

Dentro del Proyecto LEIF de Lucha contra el Empleo Irregular y el Fraude,

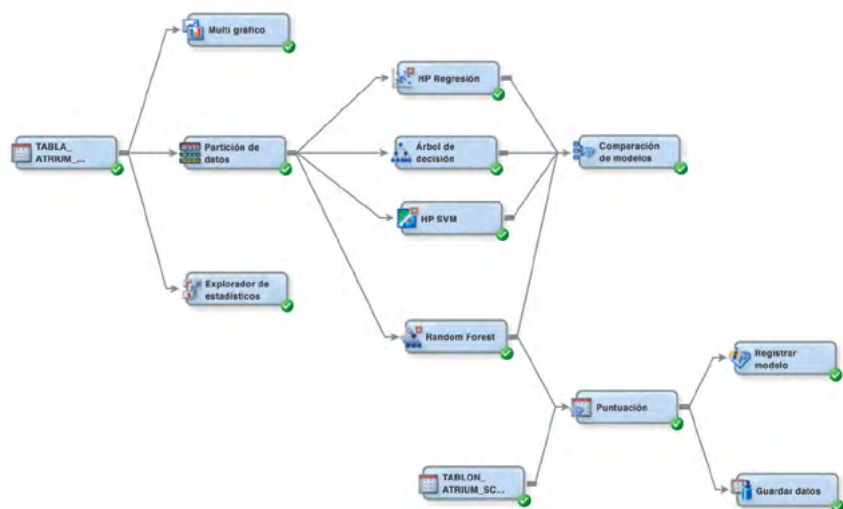
la Tesorería General de la Seguridad Social (TGSS) tiene como uno de sus objetivos la identificación de empresas que simulan una situación de actividad para perfeccionar irregularmente el derecho a cobrar con posterioridad una prestación del sistema (Desempleo, Maternidad, etc.). Los modelos analíticos crean un Score o Ranking de empresas con mayor propensión a incurrir en alguno de los supuestos contemplados. Dichos scores se incorporan a listados de alertas que se envían y dirigen la actuación de campo de las unidades de inspección a nivel provincial. Mediante modelos de visualización de Redes es posible ampliar el ámbito de estudio para la identificación de tramas de fraude<sup>4</sup>.



Figura 3. Metodología SEMMA aplicada a la investigación de fraude.

## Analítica Predictiva en la Incapacidad Temporal

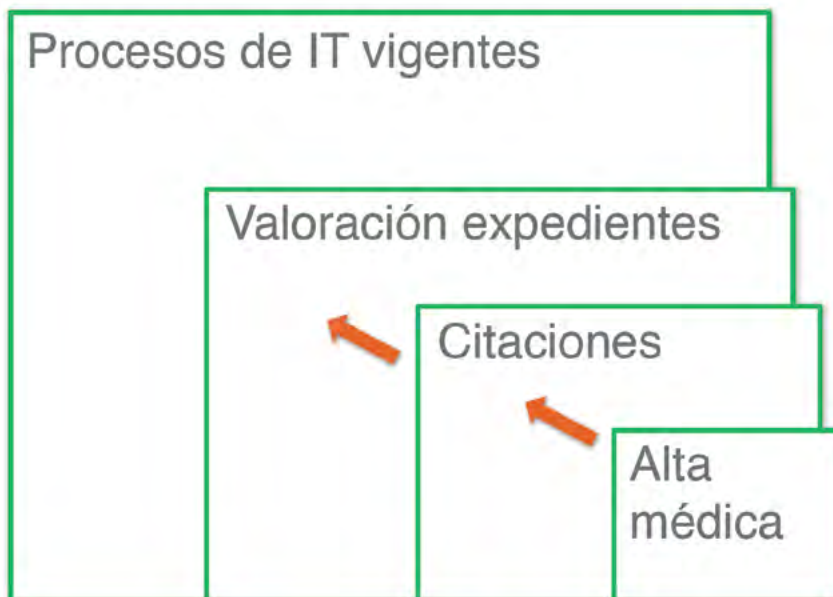
Los modelos predictivos aplicados en el ámbito de la Incapacidad Temporal (IT) persiguen incrementar la eficiencia y productividad de las Unidades de Inspección médica del Instituto Nacional de la Seguridad Social (INSS). De todos los procesos de IT existentes en un momento determinado, los inspectores médicos han de seleccionar para su valoración aquellos que estiman más propensos a resultar en alta médica tras realizar la citación del trabajador. La aplicación de modelos predictivos desarrollados con SAS permite optimizar dicho proceso de selección de manera que se incremente la tasa de altas médicas respecto del total de citaciones realizadas<sup>5</sup>. Durante la fase de modelización con SAS Miner se aplican diversas técnicas y algoritmos (Regresión logística, Árbol de decisión, Support Vector Machines, Random Forest) que compiten entre sí y se elige aquella técnica que mejor tasa de acierto ofrece<sup>6</sup>.



**Figura 6. Modelización con SAS Miner.**

En definitiva, la renovación tecnológica llevada a cabo por la GISS de su ecosistema analítico ha permitido dotar de nuevas capacidades a las Entidades Gestoras y resto de Órganos Directivos de la SESS. En el ámbito de la analítica descriptiva tradicional orientada a la toma de decisiones, las nuevas herramientas han potenciado la capacidad de

los gestores de acceder de manera autónoma a la información disponible sin depender de las distintas Unidades TIC de la GISS. Y por otro lado, la aplicación de nuevas técnicas de aprendizaje automático y análisis predictivo está permitiendo optimizar procesos de negocio, especialmente en el ámbito de la lucha contra el fraude. \*



**Figura 5. Optimización del proceso de inspección de IT.**