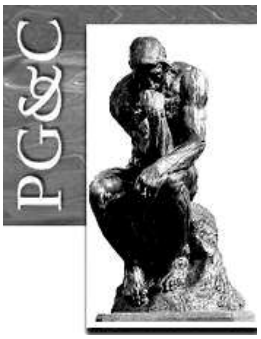

DATA MINING PARA EVALUAR EL RIESGO OPERATIVO EN PROCESOS TECNOLÓGICOS

Pedro Solana González
Doctor en Ingeniería Industrial por la Universidad de Cantabria, España
Profesor de la Universidad de Cantabria, Espanã.
E-mail: pedro.solana@unican.es

Rafael E. Bello Pérez
Doctor en Ciencias Matemáticas por la
Universidad Central "Marta Abreu" de Las Villas, Cuba.
Profesor de la Universidad Central "Marta Abreu" de Las Villas, Cuba.
E-mail: rbellop@uclv.edu.cu

Maria Matilde Garcia Lorenzo
Doctor en Ciencias Matemáticas por la
Universidad Central "Marta Abreu" de Las Villas, Cuba.
Profesor de la Universidad Central "Marta Abreu" de Las Villas, Cuba.
E-mail: mmgarcia@uclv.edu.cu

Adolfo Alberto Vanti
Doctor en Ciencias Económicas y Empresariales por la
Universidad de Deusto, España.
Profesor de la Universidade Regional Integrada do
Alto Uruguai e das Missões, Brasil
E-mail: adolfo.vanti@san.uri.br

Ivan Henrique Vey
Doctor en Ingeniería de Producción por la
Universidad Federal de Santa Catarina, Brasil.
Profesor de la Universidad Federal de Santa María, Brasil.
E-mail: ivanvey@hotmail.com

Resumen
Un riesgo operativo es un riesgo de negocio principalmente en empresas que actúan en el sector financiero. Este tipo de riesgo puede ser tratado con diferentes marcos regulatorios, los específicos de riesgo, los de seguridad y los de evaluación de procesos tecnológicos como COBIT del Instituto de Gobernanza de TI. Identificar y tratar el riesgo no siempre es tarea fácil aun con muchos estudios. En esta investigación se utiliza la metodología *Data Mining* con la técnica de *Machine Learning* basada en árboles de decisión, para analizar el proceso de *Evaluación y Gestión de Riesgos (PO9)* del dominio *Organización y Planificación* de COBIT. La base de datos se fundamenta en el grado de madurez respondido por 548 empresas en 34 procesos diferentes. Los resultados encontrados se corresponden con la jerarquía de relaciones representadas en el árbol de decisión y con la representación de otros algoritmos utilizados en un previo clasificador de transparencia de esta misma base de datos.

Palabras clave: Riesgo Operativo. Procesos Tecnológicos. Gobernanza de TI. Minería de Datos. Aprendizaje Automático.

DATA MINING TO EVALUATE OPERATIONAL RISK IN TECHNOLOGICAL PROCESSES

Abstract
An operational risk is a business risk mainly in companies that operate in the financial sector. This type of risk can be dealt with different regulatory frameworks, as risk specific, security and technological process evaluation such as COBIT from the IT Governance Institute. To Identify and treat risk is not always easy, even with many studies. In this research Data Mining methodology is used with Machine Learning technique based on decision trees, to analyze the Risk Assessment and Management (PO9) process of the Planning and Organization domain of COBIT. The database is based on the maturity level of 548 companies in 34 different processes. The results found correspond to the hierarchy of relations represented in the decision tree and with the representation of other algorithms used in a previous transparency classifier of this same database.

Keywords: *Operational Risk. Technological Processes. IT Governance. Data Mining. Machine Learning.*

1 INTRODUCIÓN

La gestión de riesgos puede ser evaluada de diferentes maneras a través de los marcos regulatorios de seguridad, los propiamente de riesgos y los de evaluación de procesos tecnológicos que buscan mejorar la calidad de la información, reducir la asimetría de la misma y aumentar la transparencia de la empresa. Los riesgos operativos (AWAD, 2018) son una preocupación continua del Pilar III de Basilea y por ello generar conocimiento a partir de los procesos tecnológicos - procesos de TI - puede representar un importante avance para el control empresarial.

En ocasiones, este mismo riesgo operativo se vuelve un riesgo de negocio estratégico porque para determinados tipos de empresas como los bancos, la información es su mayor activo. Evaluar los procesos tecnológicos a través de su madurez y saber cómo están respecto a la gestión de los datos es garantizar que se atienden los principios de Gobernanza de Tecnología de la Información. Para ello se emplean los marcos regulatorios (HEIDINGER; GATZERT, 2018) como el analizado en el estudio de *Enterprise Risk Management* (ERM) considerado en el sector bancario alemán, que actúa en este nivel de procesos tecnológicos y de riesgo.

Este estudio sin embargo ha profundizado en la familia COBIT (ALHINAI et al., 2016) respecto al proceso *PO9 - Evaluación y gestión de riesgos*; siendo los dominios de este marco de referencia: *Planear y organizar (PO)*, *Adquirir e implementar (AI)*, *Entregar y dar soporte (DS)* y *Monitorear y evaluar (ME)*.

En este trabajo se ha definido el siguiente problema de investigación: ¿Cuál es la estructura de relaciones del proceso tecnológico *PO9 - Evaluación y gestión de riesgos* de COBIT para evaluar el riesgo operativo? Para dar respuesta a este problema se ha considerado la metodología *Data Mining* (DM) asociada a técnicas de *Machine Learning* con árboles de decisión. La base de datos utilizada ha considerado 34 procesos tecnológicos en 548 empresas, teniendo éstos un análisis previo de madurez y una clasificación del nivel de transparencia (alta, media o baja).

En la actualidad la alta disponibilidad de información en soporte digital ha llevado a la necesidad e interés de convertir los datos en conocimiento, el cuál es usado en la construcción de sistemas inteligentes que ayudan a la toma de decisión basada en múltiples criterios como se detalla en los proyectos referidos por Jabbarzadeh (2018). Las técnicas actuales de análisis de datos no incluyen solamente las que usan modelos estadísticos, como se ha hecho

tradicionalmente, sino otras basadas en técnicas de la inteligencia artificial, en las que se realiza una búsqueda heurística de las relaciones existentes entre los datos.

Se han desarrollado diversas técnicas para el descubrimiento de conocimiento y definición de prioridades, entre ellas las que permiten construir modelos de predicción de variables a partir de datos conocidos, ya sea para problemas de clasificación (cuando el dominio de la variable de decisión es discreto) o en problemas de aproximación de funciones reales. Este es el caso del presente trabajo, donde se quiere analizar en el marco de gobierno de TI el riesgo operativo conforme al proceso *PO9 - Evaluación y gestión de riesgos* (MITCHELL, 1997; WITTEN; FRANK, 2005; McCUE, 2007; CHOI et al., 2017).

Se utilizaron técnicas de aprendizaje para construir un modelo computacional que permitió generar las relaciones de dependencia del riesgo con árboles de decisión, realizándose diferentes pruebas en Weka para el aprendizaje inductivo: *BFtree DecisionStump* y *NBtree*.

La selección de los árboles de decisión, en particular el algoritmo C4.5, en su versión mejorada J48, se fundamenta en dos cuestiones principales: (i) este método está entre los 10 mejores métodos de aprendizaje automático (WU et al., 2008) y (ii) es un método que se clasifica en la categoría de aprendizaje simbólico, de modo que los resultados del proceso de descubrimiento y las predicciones se realizan a partir del modelo descubierto.

Se trata de modelos interpretables, es decir, los especialistas del dominio de aplicación pueden entender esos resultados. Esta cualidad es muy apreciada en el contexto de la denominada Inteligencia Artificial Explicable - *Explainable Artificial Intelligence* (EAI) - (BIRAN; COTTON, 2017; MILLER; HOWE; SONENBERG, 2017; ABDUL et al., 2018). Se logra inferir el rasgo objetivo de nivel de transparencia en procesos tecnológicos con resultados de exactitud adecuados, además de poder explicar el porqué de la clasificación encontrada a través del árbol de decisión.

El trabajo avanza en la revisión de literatura tratando el Riesgo Operativo, el contexto de Gobernanza de TI y la evaluación de procesos tecnológicos con la familia de productos COBIT, posteriormente se expone la metodología de investigación - DM y *Machine Learning* con Árbol de Decisión -, se presentan los resultados de la investigación con el diferencial de *RandomForest*, para finalizar con las conclusiones del trabajo.

2 REFERENCIAL TEÓRICO

Para la revisión de literatura se han analizado los temas relacionados con riesgo operativo, gobernanza de tecnologías de la información y COBIT para la evaluación de procesos tecnológicos y niveles de madurez.

2.1 Riesgo operativo

El riesgo operativo está relacionado con las pérdidas no esperadas por problemas de operación del negocio internos como sistemas, personas (CROUHY; GALAI; MARK, 2008), controles, procesos, tecnologías, actividades no autorizadas, invasiones de *hackers*, mala definición de funciones, etc., pudiendo derivarse incluso de situaciones externas (CHORAFAS, 2004) considerando los riesgos cibernéticos y sus respectivos costes (ELING; WIRFS, 2019). También surge de iniciativas estratégicas de importancia, tales como el desarrollo de una nueva línea de negocio o su redefinición (CROUHY, GALAI; MARK, 2008; MERALI; PAPADOPOULOS; NADKARNI, 2012).

En el caso de las instituciones financieras, el Comité de Basilea (BCBS, 2006) define el riesgo operativo como la falta de adecuación o fallo interno resultante de procesos, personas y sistemas o también de eventos externos. Asimismo, el riesgo operativo puede ser considerado un riesgo estratégico de negocio; en un banco el fallo técnico de generación de su principal activo, la información, es directamente un riesgo estratégico que puede impactar incluso en la propia supervivencia de la empresa. Ningún cliente admitiría errores en la gestión de sus finanzas y en el corto plazo podría retirar sus fondos, la domiciliación de su nómina u otros productos financieros.

El riesgo operativo se puede tratar con un enfoque ERM (AZIZAN; SAMAD; WOON, 2011) que analiza la operación del negocio en otras actividades, como fraudes (BCBS, 2006), de una manera completa e integral para que la empresa se anticipe a las posibles "No Conformidades" (HINA; DOMINIC, 2018; KNUPLESCH; REICHERT, 2017) internas y externas, como ataques a sistemas y diversas vulnerabilidades (SINGH; MARGAM, 2018; PARK; CHAI, 2018) hasta la criptografía de datos (AWAD, 2018), la seguridad y prácticas inadecuadas en el uso de información confidencial (HASBINI; ELDAB; ALDALLAL, 2018) y de servicios (GAO; RAU; ZHANG, 2018).

Los procesos internos relacionados con planificación estratégica, arquitectura de la información, inversión en TI, gestión de recursos humanos, gestión de proyectos, adquisición de *software*, gestión de cambios, definición de niveles de servicios de TI, garantías de seguridad de información, gestión de datos, monitoreo de las TI y evaluación de los controles internos, proporcionan una adecuada gobernanza de TI, impactando directamente en las garantías de riesgos, conformidad - *compliance* - y responsabilidad - *accountability* - entre otros principios que generan mayor transparencia en las instituciones, sean públicas o privadas.

El proceso *PO9 - Evaluación y gestión de riesgos* puede ser tratado con diferentes técnicas, aunque son las técnicas más enfocadas al conocimiento y a la intangibilidad como DM y *Fuzzy Logic* (AZAR; DOLATABAD, 2019) las que se vuelven más necesarias en un contexto complejo de toma de decisiones como son las empresas actuales.

2.2 Gobernanza de tecnología de la información

Weill e Ross (2005, pp. 2-11) e ITGI (2003) definen la Gobernanza de Tecnologías de la Información (GTI) como los derechos de decisión y la matriz de responsabilidades para motivar comportamientos deseables en el uso de las TI, ampliándose también a funciones de supervisión, monitorización, control y dirección de una empresa. Este concepto continúa perfeccionándose atendiendo a la dinámica del mercado, como en Vejseli, Rossmann y Connolly (2019) que identifican 46 dimensiones de gobierno ágil en el sector de bancario.

La GTI regula en su entorno tecnológico quién es gobernado, qué y cómo se gobierna, especificando las responsabilidades, motivando el comportamiento esperado frente al uso de la tecnología en la empresa, posibilitando de esta manera una mayor alienación estratégica (LUFTMAN; BRIER, 1999; LUFTMAN, 2003) e incluso alcanzando la gobernanza de seguridad de la información y comunicaciones (GUIMARÃES; SOUZA NETO; LYRA, 2018) así como la externalización de servicios (HUYGH; DE HAES, 2018; GANTMAN; FEDOROWICZ, 2016).

La gobernanza de TI o también denominada de Gobierno de TI, atiende a la gobernanza corporativa en diferentes proyectos y estructuras (DERAKHSHAN; TURNER; MANCINI, 2019) que se concentran en la gestión y utilización de tecnologías para lograr metas de rendimiento corporativo, definiendo e implementando procesos, estructuras y mecanismos relacionados en la empresa (WEILL; ROSS, 2005; VAN GREMBERGEN; DE HAES, 2017). La GTI posibilita que los empleados ejecuten tareas utilizando tecnologías, lo que facilita converger hacia la estrategia y la competitividad de la empresa añadiendo valor a la misma (KLEIN et al.,

2019; IBGC, 2015; ISACA, 2019), aumentando la conformidad y transparencia, y reduciendo los riesgos, que es el foco principal de este trabajo.

El Instituto de Gobernanza de TI se enfoca con carácter internacional en la Gobernanza Corporativa de TI - *Enterprise Governance of IT* (EGIT) - para las direcciones empresariales sobre tecnologías de la información, contribuyendo con diferentes marcos (ISO/IEC 38500, 2015) y ampliando sus garantías con el estándar denominado COBIT que mejora a cada versión en el impacto estratégico en diferentes áreas y servicios (WAUTELET, 2018).

Esta investigación se centra en el marco COBIT de GTI considerando las percepciones de diferentes profesionales en 548 empresas. Utilizando DM como metodología cuantitativa se evalúa prioritariamente el proceso tecnológico *PO9 - Evaluación y gestión de riesgos* que de manera agrupada para los niveles de madurez permite establecer relaciones entre la calidad de la información, el aumento de transparencia y la reducción de asimetrías de la información (LEE; LEE; WANG, 2017).

2.3 COBIT

Las empresas operan y se gestionan mediante sistemas de información que están asociados a riesgos y, en este contexto, tener procesos tecnológicos bien dimensionados, estructurados y con niveles altos de madurez, como ocurre en un marco de Gobernanza de TI (ITGI, 2015), implica riesgos bajos. En esto se posiciona y actúa el estándar COBIT, que desde 1996 avanza para posibilitar aplicaciones más robustas (ALKHALDI; HAMMAMI; UDDIN, 2017) en las áreas de gestión, tecnología y buen gobierno tecnológico.

COBIT es un marco de gestión de TI que contribuye al desarrollo, organización e implementación de estrategias basadas en información y gobernanza, donde la reducción de riesgos se asocia a los principios de gobierno corporativo (OCDE, 2015) a través de la gobernanza tecnológica (ISACA, 2012). Este importante marco internacional es utilizado en las más diversas empresas y aplicaciones, como en la comunicación y control en proyectos contratados de sistemas de información (GANTMAN; FEDOROWICZ, 2016).

El marco de referencia COBIT cuenta con 4 dominios: PO - Planear y Organizar, AI - Adquirir e Implementar, DS - Entregar y dar Soporte y ME - Monitorear y Evaluar; además de con 34 procesos y 220 objetivos de control. Asimismo, desde 2012 integra objetivos de TI y habilitadores de COBIT que consideran información, procesos y estructuras organizacionales, incluyendo también la evaluación de procesos tecnológicos y riesgos (DEBRECENY; GRAY, 2013) a través del análisis de sus niveles de madurez (TARMUJI et al., 2017; MURILLO PAEZ; TINOCO DIAZ; CARRERA NARANJO, 2019) en un contexto de auditoría de cumplimento (ISACA, 2012). Asimismo, destacan los trabajos de Aguiar et al., (2018) profundizan en los modelos de madurez para la gestión de incidentes considerando múltiples marcos, incluyendo ITIL y CMMI-SCV. En este ambiente tecnológico es habitual integrar diferentes técnicas para alcanzar los objetivos definidos, como ha pasado en procesos de gobierno y administración de TI.

De este modo los estándares van evolucionando en versiones que gradualmente reconfiguran y sofistican los dominios - *Alinear, planificar y organizar (APO)*, *Construir, adquirir e implementar (BAI)*, *Entregar, dar servicio y soporte (DSS)* y *Supervisar, evaluar y valorar (MEA)* - que soportan la gobernanza de TI a través de una metodología que asegura que el área de TI esté alineada con los propósitos de la compañía, habilitando el negocio para maximizar los beneficios, que los recursos de TI son utilizados responsablemente y los riesgos de TI son gestionados apropiadamente. Se introduce en la evaluación de procesos para analizar la madurez la norma ISO/IEC 15504, que perfecciona la anterior basada en CMMI y con alcance en 2019 se referencia el uso de la ISO/IEC 33000.

COBIT en su versión 2019 (ISACA, 2019) considera una estructura para la Gobernanza y Gestión Corporativa de TI para toda la organización, convergiendo con una tecnología

corporativa adecuada creando estrategias de gobernanza más flexibles y colaborativas (*online*) que atienden a las tecnologías digitales y también de seguridad de información, presentando un Core Model como actualización del Modelo de Referencia de Procesos. Asimismo, actúa en principios que añaden áreas de enfoque, perfecciona el modelo de madurez añadiendo 3 nuevos objetivos que facilitan aún más la gestión y los factores de diseño, simplificando su implantación en las empresas con mayor eficiencia operacional y flexibilidad, para posibilitar nuevas soluciones de gobernanza.

Incorpora el modelo ISO/IEC 33000 para evaluación de la madurez y capacidad en un sistema más completo de gobernanza con componentes, una estructura de gobierno más dinámica respecto a los cambios tecnológicos y también con mayor definición de actividades y necesidades de empresas que actúan en ambientes más complejos de decisión y de seguridad informacional. El marco de referencia COBIT mantiene una compatibilidad en su evolución (ITGI, 2015; AGUIAR et al., 2018) para proporcionar a las empresas la flexibilidad de desarrollar soluciones prácticas de gobierno, contribuyendo a reducir la brecha entre TI y gestión.

Es importante que se perfeccionen continuamente los enfoques y técnicas para evaluar procesos, como con la utilización de métodos de decisión multicriterio, el análisis de datos de manera amplia con minería de datos o con el análisis de bases de datos que muestre en un nivel de detalle lo que está pasando en la empresa. Por ello se establece como metodología de este trabajo un enfoque de *Design Science* en una configuración de *Data Mining* en la que la máquina aprende.

3 METODOLOGÍA

En el modelo computacional para *PO9 - Evaluación y gestión de riesgos* conforme a procesos de TI de COBIT, se han utilizado técnicas de DM (CHOI et al., 2017) con la implementación mediante la herramienta *Weka* (2019) de código abierto, ampliamente utilizada por investigadores en todo el mundo. Este enfoque *Data Science* está definido como un conjunto de principios fundamentales que conducen a la extracción de conocimiento a partir de los datos y que se asocian con *Machine Learning* (McCUE, 2007), de modo que la máquina aprende con los datos e incorpora nuevo conocimiento a la misma.

De esta manera, *Data Science* engloba los principios, procesos y técnicas para comprender los fenómenos a través del análisis de datos, la previsión y el aprendizaje automático (WAN et al., 2016). En esta investigación ha sido aplicado este enfoque a 548 empresas considerando los procesos relacionados con la planificación de TI, pasando por la gestión de proyectos, riesgos, seguridad, contratación de servicios a terceros, la monitorización de la gestión y TI, para finalmente controlar la adecuada Gobernanza de TI (GTI).

La minería de datos combina técnicas estadísticas, inteligencia artificial y reglas de negocio (FARAZZMANESH; HOSSEINI, 2017) para encontrar patrones que las técnicas usuales como *Business Intelligence* (BI) o una búsqueda de correlaciones normalmente no encuentran.

Los procesos tecnológicos basados en la referencia a la evolución de COBIT han sido evaluados según su grado de madurez en 6 niveles:

- 0 - *Inexistente*: En este nivel hay una absoluta falta de proceso. La organización no tiene conocimiento referente a las implicaciones que la falta de proceso sistematizado puede generar.
- 1 - *Inicial*: En este nivel los procesos son esporádicos y desorganizados, no existe documentación ni control.
- 2 - *Repetitivo*: En este nivel los procesos siguen un patrón de regularidad, con alta dependencia del conocimiento de los individuos.
- 3 - *Definido*: En este nivel los procedimientos están establecidos y se cumplen. Inicio del uso de indicadores para el control.
- 4 - *Gestionado*: En este nivel los procesos están integrados y alineados. Las metas y planes están basados en datos e indicadores consistentes.
- 5 - *Optimizado*: Se consiguen y automatizan buenas prácticas con base en los resultados de la mejora continua.

De esta manera, se ha medido el grado de madurez consolidado en 34 procesos de la organización evaluando el nivel de riesgo operativo (PO9) con otros procesos de COBIT. La investigación realizada se extendió a empresas en un estudio longitudinal que viene recogiendo datos desde abril de 2012, por lo que se presenta de manera robusta generando una continuidad en la investigación científica que alcanza las versiones del marco de referencia COBIT que sigue cambiando e incorporando los enfoques anteriores.

Inicialmente se han utilizado técnicas de descubrimiento (MITCHELL, 1997; McCUE, 2007) para construir un modelo computacional F que calcula también el nivel de transparencia (NT) de una organización (O) y que contextualiza el riesgo considerando los procesos tecnológicos con el enfoque de madurez de COBIT. Se ha encontrado la estructura jerarquizada con la que se pueden visualizar las dependencias entre los procesos que mantienen la gestión de la información en las empresas en un proceso de selección de atributos conforme al presentado en la Figura 1.

Esa jerarquía se expresa en forma de árbol de decisión, donde los nodos del árbol representan atributos o rasgos del problema, y de cada nodo salen tantos enlaces como valores tiene el dominio del rasgo. En la raíz del árbol se sitúa el rasgo con mayor poder de separación de los casos; en el caso del método empleado ese valor se corresponde con la reducción de la entropía; es decir, se calcula la entropía del conjunto de casos usando la expresión (1) y luego se calcula para el rasgo en cuanto que este reduce la entropía si se utiliza para separar el conjunto de datos y subconjuntos que tienen igual valor para ese rasgo, empleando la expresión (2). El proceso se continúa por cada rama del árbol hasta encontrar subconjuntos que pertenecen a la misma clase.

(1)

$$Entropia(S) = \sum_{i=1}^{c} - p_i \log_2 p_i \qquad siendo \quad 0 \log_2 0 = 0$$

$$p_i = \frac{|Si|}{|S|}$$

Donde S el conjunto de ejemplos, los ejemplos pertenecen a c clases y Si denota el subconjunto de ejemplos que pertenecen a la clase i. La medida de efectividad de un atributo en clasificar el conjunto de ejemplos se define a partir de cómo reduce la entropía del conjunto. Sea esta medida denotada por Gain (S, A), que indica la reducción esperada de la entropía causada por el conocimiento del valor del atributo A.

(2)

$$Gain(S, A) = Entropia(S) - \sum_{v \in Valores(A)} \frac{|Sv|}{|S|} * Entropia(Sv)$$

Donde Valores (A) es el conjunto de valores posibles (dominio) del atributo A y Sv es el subconjunto de S para el cual el atributo A tiene valor v, es decir, Sv = {s∈S | A(s) = v}. En la Figura 2 se puede apreciar como el atributo de mayor jerarquía *a MO1 - Monitorización y evaluación del desempeño de TI* para menores o iguales que 3, le siguen en orden de importancia *MO2 - Monitorización y evaluación del control interno*, *DS10 – Gestión de los problemas* y *PO9 - Evaluación y gestión de los riesgos de TI* cuando MO2 es mayor que 1.

Cuando el valor de *MO2 - Monitorización y evaluación del control interno* es 0 o 1 aparecen otros atributos a analizar, como *PO5 - Gestión de la inversión en TI* y *PO8 – Gestión de la calidad*, pero en la mayoría de los casos el nivel de transparencia es bajo.

Se ha utilizado el aprendizaje inductivo con árboles de decisión en su variante implementada en *Weka* (WITTEN; FRANK, 2005; WEKA, 2019), para comprender el riesgo operativo de acuerdo con los profesionales de empresas que contestaron al instrumento de recogida de datos vía *Google Drive.* De este modo se han considerado bajo sus percepciones los grados de madurez consolidados de los procesos en la escala de 0 a 5, anteriormente indicada. El enfoque de árbol de decisión con algoritmo J48 ha posibilitado la calidad del conocimiento descubierto (HUANG; LING, 2005). Para ello se seleccionaron los atributos principales de los procesos más relevantes para posteriormente tratar el árbol de decisión conforme con la ejemplificación que se visualiza en la investigación. La Figura 1 muestra la selección de atributos que prioriza determinados procesos.

Figura 1 – Selección de atributos

```
=== Run information ===
Evaluator:   weka.attributeSelection.CfsSubsetEval -P 1 -E 1
Search:      weka.attributeSelection.BestFirst -D 1 -N 5
Relation:    Figura 1 cobitnuevo500-weka.filters.unsupervised.attribute.Remove-R1,3
Instances:   547
Attributes:  36
             .....
Evaluation mode:    evaluate on all training data

== Attribute Selection on all input data ===d
Search Method:
        Best first.
        Start set: no attributes
        Search direction: forward
        Stale search after 5 node expansions
        Total number of subsets evaluated: 602
        Merit of best subset found:    0.59
Attribute Subset Evaluator (supervised, Class (nominal): 36 ESCALA NÍVEIS DE TRANSPARÊNCIA):
        CFS Subset Evaluator
        Including locally predictive attributes
                Selected attributes:
3,4,5,6,7,8,9,10,11,12,13,14,15,17,18,19,20,21,23,24,25,26,27,28,29,30,31,32,33,34,35: 31
            PO2 - Define la arquitectura de la información.
            PO3 - Determina la dirección tecnológica.
            PO4 - Define procesos, organización y relaciones de TI.
            PO5 - Gestiona la inversión en TI.
            PO6 - Comunica las metas y las directrices gerenciales.
            PO7 - Gestiona los recursos humanos de TI.
            PO8 - Gestiona la calidad.
            PO9 - Evalúa y gestiona los riesgos de TI.
            PO10 - Gestiona los proyectos.
            AI1 - Identifica soluciones automatizadas.
            AI2 - Adquiere y mantiene el software aplicativo.
            AI3 - Adquiere y mantiene la arquitectura tecnológica.
            AI4 - Desarrolla y mantiene procedimientos de TI.
            AI6 - Gestiona los cambios.
            AI7 - Instala e certifica soluciones y cambios.
            DS1 - Define y gestiona niveles de servicio.
            DS2 - Gestiona los servicios de terceros.
            DS5 - Garantiza la seguridad de los sistemas.
            DS6 - Identifica y asigna costes.
            DS7 - Educa y entrena a los usuarios.
            DS8 - Gestiona la central de servicios y los incidentes.
            DS9 - Gestiona la configuración.
            DS10 - Gestiona los problemas.
            DS11 - Gestiona los datos.
            DS12 - Gestiona la infraestructura.
            DS13 - Gestiona las operaciones.
            MO1 - Monitoriza y evalúa el desempeño de TI.
            MO2 - Monitoriza y evalúa el controle interno.
            MO3 - Asegura la conformidad con los requisitos externos.
            MO4 - Proporciona el gobierno de TI.
```

Fuente: investigación propia

A partir de la selección de los atributos más importantes para alcanzar el problema científico investigado el árbol empieza a ser estructurado optando por diferentes caminos que permiten comprender mejor sus jerarquías. Se evidencian así de manera analítica y visual las dependencias entre procesos tecnológicos de todo un grupo de empresas, y se determina en qué procesos prioritarios se deben utilizar recursos financieros. De esta manera, se priorizan a continuación los resultados generados a través de diferentes análisis.

4 RESULTADOS

Una vez realizada la selección de atributos se construyó el árbol de decisión para predecir inicialmente el nivel de transparencia, alcanzándose más del 96% de clasificaciones correctas y un error cuadrático medio igual a 0.1 cuando se utiliza como muestra todo el conjunto de entrenamiento, siendo la clase de peores resultados la media (Figura 3).

En la Figura 2 se muestra un pequeño fragmento del árbol de decisión donde puede apreciarse el nivel de incidencia de *PO9 - Evaluación y gestión de riesgos* para obtener esa clasificación. Se puede llegar a una clasificación del nivel de transparencia conocido, tal que si *MO2 - Monitora y evalúa el control interno* >1 y *DS10 - Gestiona los problemas* <=1 y *PO10 - Gestiona los proyectos* <=3 con nivel de transparencia bajo, entonces *PO9* será menor o igual al nivel de madurez 2 (*PO9*<=2).

Por lo tanto, es posible comprender la posición de *PO9* y de otros procesos dentro de su contexto de relaciones en el árbol de decisión. De este modo, entendiendo las relaciones directas existentes es posible invertir en procesos tecnológicos específicos y consecuentemente aumentar el grado de madurez que más interese a la empresa, como es el caso del proceso *PO9* estudiado en esta investigación.

Figura 2 – Resultados para árbol de decisión J48

```
J48 pruned tree
------------------
MO1 - Monitoriza y evalúa el desempeño de TI. <= 3
|   MO2 - Monitoriza y evalúa el control interno. <= 1
|   |   PO5 - Gestiona la inversión en TI. <= 3
|   |   |   PO8 - Gestiona la calidad. <= 3: BAJA (127.41/1.0)
|   |   |   PO8 - Gestiona la calidad. > 3
|   |   |   |   PO1 - Define la planificación estratégica de TI. <= 2: BAJA (7.0)
|   |   |   |   PO1 - Define la planificación estratégica de TI. > 2: MEDIA (2.0)
|   |   PO5 - Gestiona la inversión en TI. > 3
|   |   |   PO4 - Define procesos, organización y relaciones de TI. <= 2: BAJA (6.0)
|   |   |   PO4 - Define procesos, organización y relaciones de TI. > 2
|   |   |   |   DS6 - Identifica y asigna costes. <= 2: BAJA (2.0)
|   |   |   |   DS6 - Identifica y asigna costes. > 2: MEDIA (10.0/1.0)
|   MO2 - Monitora y evalúa el control interno. > 1
|   |   DS10 - Gestiona los problemas. <= 1
|   |   |   PO9 - Evalúa y gestiona los riesgos de TI. <= 2
|   |   |   |   PO10 - Gestiona los proyectos. <= 3: BAJA (36.15/1.0)
|   |   |   |   PO10 - Gestiona los proyectos. > 3: MEDIA (2.0)
|   |   |   PO9 - Evalúa y gestiona los riesgos de TI. > 2
|   |   |   |   PO5 - Gestiona la inversión en TI. <= 1: BAJA (3.0)
|   |   |   |   PO5 - Gestiona la inversión en TI. > 1
|   |   |   |   |   AI6 - Gestiona los cambios. <= 1: BAJA (2.0)
|   |   |   |   |   AI6 - Gestiona los cambios. > 1: MEDIA (13.0/1.0)
```

Fuente: investigación propia

La Figura 2 muestra un fragmento del árbol construido y el nivel de importancia de los rasgos en función de la clasificación. De este modo, el rasgo que más discrimina el desorden resulta ser *MO1 - Monitoriza y evalúa el desempeño de TI* (ubicado en la raíz del árbol) y dependiendo del valor de esta variable se originan otros niveles en el árbol de decisión.

El árbol de decisión estructura de manera significativa y comprensible lo que está ocurriendo con los datos, considerando los procesos y sus relaciones. También se pueden comprender los niveles de transparencia conjugados con cada proceso estudiado y verificar cómo se comportan los demás procesos respecto al que se ha fijado. Esta técnica permite hacer diversos análisis de la posición de cada proceso y sus relaciones dentro del árbol de decisión, considerando el nivel de transparencia Alto, Medio o Bajo.

En la Figura 3 se presentan los resultados de la evaluación para todo el conjunto de entrenamiento en *Weka*. Éste permite entrenar la base de datos con partes, fragmentando y agrupando datos de diferentes maneras.

Figura 3 – Resultados de la evaluación para todo el conjunto de entrenamiento

```
=== Summary ===
Correctly Classified Instances         529               96.7093 %
Incorrectly Classified Instances        18                3.2907 %
Kappa statistic                          0.9489
Mean absolute error                      0.0401
Root mean squared error                  0.1397
Relative absolute error                  9.3502 %
Root relative squared error             30.1643 %
Total Number of Instances              547

=== Detailed Accuracy By Class ===
               TP Rate   FP Rate   Precision   Recall  F-Measure  MCC      ROC Area  PRC Area  Class
                 0,972    0,012    0,981     0,972    0,976      0,961    0,995     0,989     Baja
                 0,955    0,025    0,963     0,955    0,959      0,932    0,985     0,976     Media
                 0,982    0,014    0,949     0,982    0,966      0,956    0,995     0,967     Alta
Weighted Avg.    0,967    0,017    0,967     0,967    0,967      0,948    0,991     0,979
```

Fuente: investigación propia

Utilizando otras técnicas de construcción de árboles de decisión, el árbol generado es similar al representado pero los resultados de la evaluación son inferiores, solo mejorados por *Random Forest* que funciona como combinación de árboles de decisión y, por lo tanto, no visualiza el árbol. La Tabla 1 muestra los resultados alcanzados con varios métodos de aprendizaje supervisado del tipo árbol de decisión. Se aplicaron diferentes técnicas - *NBTree*, *BFTree*, *DecisionStump* y *RandomForest* - considerando el rendimiento de cada una.

Tabla 1 – Resultados de aprendizaje inductivo con árbol de decisión

Métodos empleados	% de clasificaciones correctas	Error cuadrático medio
NBTree	88.11	0.27
BFTree	87.02	0.26
DecisionStump	65.44	0.39
RandomForest	99.6	0.08

Fuente: investigación propia

Los resultados de aprendizaje inductivo señalan que *RandomForest* sería la técnica más adecuada para la construcción del árbol de decisión.

5 CONCLUSIONES

Este trabajo ha profundizado en la evaluación de riesgos en el marco de control de procesos tecnológicos bajo el estándar COBIT del Instituto de Gobernanza de TI, marco muy difundido internacionalmente y utilizado en grandes empresas, especialmente en las que actúan en el mercado financiero como son las entidades financieras y Bancos Centrales.

Se ha podido construir un clasificador inteligente de transparencia en el que el sistema aprende con los datos (aprendizaje de máquina) y genera conocimiento sistematizado para posicionar el riesgo operativo, la asimetría y calidad de la información, entre otros aspectos relevantes de la gestión empresarial.

La investigación se ha centrado en el proceso *PO9 - Evaluación y gestión de riesgos*, para comprender su relación con otros procesos operativos de TI a partir del árbol de decisión generado con el algoritmo J48. Se pueden simular diversas situaciones en la base de datos para comprender por ejemplo donde o en qué proceso se necesita una mayor inversión de recursos para alcanzar niveles de madurez altos.

De este modo, es posible por ejemplo saber en qué procesos deberá invertir la empresa para que alcanzar un nivel de madurez 4 (gestionado) en relación a PO9. Para ello, en este estudio se han utilizado diferentes técnicas con árboles de decisión *NBTree*, *BFTree*, *DecisionStump* y *RandomForest*, siendo esta última la que ha presentado mejores resultados.

Se utilizó la herramienta computacional *Weka* que permitió con técnicas de *Machine Learning* evaluar jerárquicamente las relaciones entre distintos procesos corporativos y las posibles inversiones en recursos tecnológicos. Como consecuencia, se ha podido contribuir a que las empresas comprendan la forma de reducir o gestionar mejor los riesgos organizacionales, perfeccionando una estructura de gobierno tecnológico que reduzca las asimetrías de la información mejorando su calidad.

Para futuras investigaciones será necesario direccionar esfuerzos a ampliar el enfoque de manera comparativa internacional, considerando compañías en diferentes regiones con distintas culturas empresariales y en entidades privadas y públicas. Asimismo, con un modelo de *Machine Learning* como el aquí presentado es posible anticipar evaluaciones y con esto direccionar mejor las inversiones organizacionales en procesos tecnológicos que generen mayor garantía en la rendición de cuentas.

Para finalizar, cabe señalar que con los avances en esta línea de investigación y a través de este tipo de estudios será posible contribuir a una cultura de mejora de los procesos empresariales, considerando de manera más sistemática los requisitos de auditoría alineados con el aumento de las garantías de gobernanza corporativa y de TI en las empresas.

REFERENCIAS

ABDUL, A.; VERMEULEN, J.; WANG, D.; LIM, B.; KANKANHALLI, M. Trends and trajectories for explainable, accountable and intelligible systems: An HCI research agenda. In: **Proceedings of the SIGCHI Conference on Human Factors in Computing Systems**, 2018.

AGUIAR, J.; PEREIRA, R.; VASCONCELOS, J.; BIANCHI, I. An overlap less incident management maturity model for multi-framework assessment (ITIL, COBIT, CMMI-SVC). **Interdisciplinary Journal of Information, Knowledge, and Management**, n. 13, p. 137-163, 2018.

ALHINAI, Y.; AL-BADI, A.; AL-HARTHI, I.; AL-SALTI, Z. Rethinking IT-governance: Analytics review of IT governance for social media based on the COBIT standard. **International Journal of Services, Economics and Management**, v. 7, n. 2-4, p. 124-153, 2016.

ALKHALDI, F. M.; HAMMAMI, S. M.; UDDIN, M. A. Understating value characteristics toward a robust IT governance application in private organizations using COBIT framework. **International Journal of Engineering Business Management**, n. 9, p. 1-8, 2017.

AZAR, A.; DOLATABAD, K. A method for modeling operational risk with fuzzy cognitive maps and Bayesian belief networks. **Expert Systems with Applications**, v. 115, p. 607-617, 2019.

AZIZAN, N. A.; SAMAD, M. F. A.; WOON, L. F. A strategic framework for value enhancing enterprise risk management. **Journal of Global Business and Economics**, v. 2, n. 1, 23-47, 2011.

AWAD, A. I. Introduction to information security foundations and applications. In: **Information security:** Foundations, technologies and applications. The Institution of Engineering and Technology, p. 3-11, 2018.

BCBS. **International convergence of capital measurement and capital standards: A revised framework**. Switzerland: Basel Committee on Banking Supervision, 2006. Disponible en: https://bit.ly/2h2ocBt. Acceso: 30 octubre 2018.

BIRAN, O.; COTTON, C. Explanation and justification in machine learning: A survey. In: **Proceedings of the IJCAI Workshop on Explainable Artificial Intelligence**, 2017.

CHOI, J.; KIM, B.; HAHN, H.; PARK, H.; JEONG, Y.; YOO, J.; JEONG, M. K. Data mining-based variable assessment methodology for evaluating the contribution of knowledge services of a public research institute to business performance of firms. **Expert Systems with Applications**, n. 84, p. 37-48, 2017.

CHORAFAS, D. N. **Operational risk control with Basel II: Basic principles and capital requirements**. Oxford: Elsevier, 2004.

CROUHY, M.; GALAI, D.; MARK, R. **Fundamentos da gestão de risco**. Rio de Janeiro: Qualitymark, 2001.

CROUHY, M.; GALAI, D.; MARK, R. **Risk management**. New York: McGraw-Hill, 2008.

DEBRECENY, R.; GRAY, G. IT governance and process maturity: A multinational field study. **Journal of Information Systems**, v. 27, n. 1, p. 157-188, 2013. DOI: 10.2308/isys-50418.

DERAKHSHAN, R.; TURNER, R.; MANCINI, M. Project governance and stakeholders: A literature review. **International Journal of Project Management**, v. 37, n. 1, p. 98-116, 2019.

ELING, M.; WIRFS, J. What are the actual costs of cyber risk events? **European Journal of Operational Research**, v. 272, n. 3, p. 1109-1119, 2019.

FARAZZMANESH, F.; HOSSEINI, M. Analysis of business customers' value network using data mining techniques. **Journal of Information Systems and Telecommunication**, v. 5, n. 3, p. 162-171, 2017.

GANTMAN, S.; FEDOROWICZ, J. Communication and control in outsourced IS development projects: Mapping to COBIT domains. **International Journal of Accounting Information Systems**, n. 21, p. 63-83, 2016.

GAO, F.; RAU, P. L. P.; ZHANG, Y. Perceived mobile information security and adoption of mobile payment services in China. In: **Mobile commerce:** Concepts, methodologies, tools, and applications. IGI Global, p. 1179-1198, 2018.

GUIMARÃES, R.; SOUZA NETO, J.; LYRA, M. Modelo de governança de segurança da informação para a administração pública federal. **Perspectivas em Gestão & Conhecimento**, João Pessoa, v. 8, n. 3, p. 90-109, 2018. DOI: 10.21714/2236-417X2018v8n3p90.

HASBINI, M. A.; ELDABI, T.; ALDALLAL, A. Investigating the information security management role in smart city organisations. **World Journal of Entrepreneurship, Management and Sustainable Development**, v. 14, n. 1, p. 86-98, 2018.

HEIDINGER, D.; GATZERT, N. Awareness, determinants and value of reputation risk management: Empirical evidence from the banking and insurance industry. **Journal of Banking and Finance**, v. 91, p. 106-118, 2018. DOI: 10.1016/j.jbankfin.2018.04.004.

HINA, S.; DOMINIC, P. D. Information security policies' compliance: a perspective for higher education institutions. **Journal of Computer Information Systems**, p. 1-11, 2018. DOI: 10.1080/08874417.2018.1432996.

HUYGH, T.; De HAES, S. Answering key global IT management concerns through IT governance and management processes: A COBIT 5 view. In: **Proceedings of the 51st Hawaii International Conference on System Sciences**, p. 5355-5344, 2018. URI: http://hdl.handle.net/10125/50554.

IBGC. **Código das melhores práticas de governança corporativa**. 5. ed. São Paulo, Instituto Brasileiro de Governança Corporativa, 2015.

ISACA. **COBIT 5: Modelo corporativo para governança e gestão de TI da organização**. Illinois, USA: ISACA, 2012.

ISACA. **COBIT 2019 Framework:** governance and management objectives**.** Illinois, USA: ISACA, 2019.

ISO/IEC 15504. **Software process improvement capability determination**. International Organization for Standardization, 2019.

ISO/IEC 33000. **Software processes evaluation**. International Organization for Standardization, 2019.

ISO/IEC 38500. **Information technology - Governance of IT for the organization**. International Organization for Standardization, 2015.

ITGI. **Board briefing on IT governance**. 2nd Edition. Illinois, USA: IT Governance Institute, 2015. Disponible en: https://bit.ly/2cOBheB. Acceso: 8 noviembre 2018.

JABBARZADEH, A. Application of the AHP and TOPSIS in project management. **Journal of Project Management**, v. 3, n. 2, p. 125-130, 2018. DOI: 10.5267/j.jpm.2018.1.001.

KLEIN, P.; MAHONEY, J.; McGAHAN, A.; PITELIS, C. Organizational governance adaptation: who is in, who is out, and who get what. **Academy of Management Review**, v. 44, n. 1, p. 6-27, 2019.

KNUPLESCH, D.; REICHERT, M. A visual language for modeling multiple perspectives of business process compliance rules. **Software & Systems Modeling**, v. 16, n. 3, p. 715-736, 2017.

LEE, H.; LEE, H-L.; WANG, C-C. Engagement partner specialization and corporate disclosure transparency. **International Journal of Accounting**, v. 52, n. 4, p. 354-369, 2017. DOI: 10.1016/j.intacc.2017.10.001.

LUFTMAN, J. Assessing IT/Business alignment. **Information Systems Management**, v. 20, n. 4, p. 9-15, 2003. DOI: 10.1201/1078/43647.20.4.20030901/77287.2

LUFTMAN, J.; BRIER, T. Achieving and sustaining business-IT Alignment**. California Management Review**, v. 42, n. 1, p. 109-122, 1999.

McCUE, C. **Data mining and predictive analysis:** Intelligence gathering and crime analysis. Oxford: Elsevier, 2007.

MERALI, Y.; PAPADOPOULOS, T.; NADKARNI, T. Information systems strategy: Past, present, future? **Journal of Strategic Information Systems**, n. 21, p. 125-153, 2012.

MILLER, T.; HOWE, P.; SONENBERG, L. Explainable AI: Beware of inmates running the asylum. In: **WORKSHOP ON WORKSHOP ON EXPLAINABLE ARTIFICIAL INTELLIGENCE**, 2017. Proceedings […], 2017.

MITCHELL, T. **Machine learning**. New York: McGraw-Hill, 1997.

MURILLO PAEZ, D.; TINOCO DIAZ, E.; CARRERA NARANJO, C. La evaluación de las tecnologías de la información usando Cobit Assurance para una auditoría de cumplimiento, **Revista Espacios**, v. 40, n. 3, p. 13-22, 2019.

OECD. **G20/OECD principles of corporate governance**. Paris, France: OECD Publishing, 2015. DOI: 10.1787/9789264236882-en.

PARK, M.; CHAI, S. Internalization of information security policy and information security practice: A comparison with compliance. In: HAWAII INTERNATIONAL CONFERENCE ON SYSTEM SCIENCES, 51., 2018. **Proceedings** […]. Association for Information Systems, 2018.

SINGH, V.; MARGAM, M. Information security measures of libraries of central universities of Delhi: A study. **DESIDOC Journal of Library & Information Technology**, v. 38, n. 2, p. 102-109, 2018. DOI: 10.14429/djlit.38.2.11879

TARMUJI, A.; SETIADI, T.; HANDAYANINGSIH, S.; LESTARI, J. Development of a customer relationship management model based on maturity level of Cobit 4.1: Case study of the cooperative section at department of industry, trade, cooperative, and small-medium enterprises, Yogyakarta province. **Asia-Pacific Journal of Science and Technology**, v. 22, n. 2, p. 1-6, 2017.

VAN GREMBERGEN, W.; DE HAES, S. Introduction to the Minitrack IT governance and its mechanism. In: HAWAII INTERNATIONAL CONFERENCE ON SYSTEM SCIENCES, 50., 2017, **Proceedings** […] p. 5162-5163, 2017.

VEJSELI, S.; ROSSMANN, A.; CONNOLLY, T. IT governance and its agile dimensions: Exploratory research in the banking sector. In: HAWAII INTERNATIONAL CONFERENCE ON SYSTEM SCIENCES, 52., **Proceedings** […]. p. 6209-6218, 2019. URI: https://hdl.handle.net/10125/60055.

WAN, J.; YUE, Z. L.; YANG, D.-H.; YU, Z.; JIAO, L.; ZHI, L.; LIU, J. Predicting non performing loan of business bank with data mining techniques. **International Journal of Database Theory and Application**, v. 9, n. 12, p. 23-34, 2016.

WAUTELET, Y. A model-driven IT governance process based on the strategic impact evaluation of services. **Journal of Systems and Software**, v. 149, p. 462-475, 2018. DOI: 10.1016/j.jss.2018.12.024.

WEILL, P.; ROSS, J. W. **IT governance:** how top performers manage IT decision rights. USA: Harvard Business School Press, 2005.

WEKA. Waitako environment for knowledge analysis. Disponible en: http://www.cs.waikato.ac.nz/ml/weka. Acceso: 12 febrero de 2019.

WITTEN, I.; FRANK, E. **Data minig:** practical machine learning tools and techniques. 2. nd Edition. San Francisco: Elsevier, 2005.

WU, X.; KUMAR, V.; QUINLAN, J. R.; GHOSH, J.; YANG, Q.; MOTODA, H. et al. Top 10 algorithms in data mining. **Knowledge Information System**, v. 14, n. 1, p. 1-37, 2008. DOI: 10.1007/s10115-007-0114-2.

Artigo recebido em 08/11/2018 e aceito para publicação em 05/06/2019