

Medición de niveles de madurez tecnológica aplicando técnicas de minería de datos

Cobo, Angel¹ acobo@unican.es

Rocha, Eliana Rocío² rochar@unican.es

Vanti, Adolfo Alberto³ avanti@unisinis.br

Campos, Rafael Herden³ rcamposrs@yahoo.com.br

¹ *Departamento de Matemática Aplicada y Ciencias de la Computación*

² *Departamento de Administración de Empresas*

Universidad de Cantabria (España)

³ *Departamento de Ciencias Contables*

Universidade do Vale do Rio dos Sinos – UNISINOS (Brasil)

RESUMEN

En este trabajo se utilizan técnicas de minería de datos para el análisis de los niveles de madurez de procesos relativos a la gobernanza de las tecnologías de la información (TI) en las organizaciones. Una de las necesidades básicas de cualquier organización es precisamente entender el estado de sus propios sistemas de TI y decidir los niveles de gestión y control a aplicar. Tomando como referencia el marco de trabajo COBIT, que define 34 procesos diferentes organizados en 4 dominios e incluyendo pautas de valoración de los niveles de madurez de cada proceso, en este trabajo se ha tratado de extraer conocimiento mediante técnicas de clustering, selección de atributos y búsqueda de reglas de asociación entre procesos. Como caso práctico se ha analizado una muestra de 78 empresas brasileñas de diferentes sectores que operan en el estado de Rio Grande do Sul, realizándose un diagnóstico inicial de los niveles de madurez de los procesos de COBIT y aplicando las técnicas de minería de datos para identificar grupos de empresas que responden a un mismo modelo de gobernanza de TI y analizando correlaciones entre procesos.

ABSTRACT

In this paper we use data mining techniques to analyze the maturity levels of processes concerning the governance of Information Technology (IT) in organizations. Understanding the state of the IT systems and deciding the management and control processes are basic needs of any organization. In the research we use COBIT, a framework which defines 34 different processes organized into 4 domains and including guidelines for assessing the maturity levels of each process. In order to extract new knowledge, clustering techniques, feature selection strategies and association rules are used. As a practical case we analyzed a sample of 78 Brazilian companies from different sectors operating in the state of Rio Grande do Sul. We performed an initial assessment of the maturity levels of the processes of COBIT and applied data mining techniques to identify groups of firms responding to the same IT governance model. Correlations between IT processes were also obtained using association rules.

Palabras claves:

Gobernanza de TI; minería de datos; COBIT; madurez

Keywords:

IT Governance; data mining; COBIT; maturity

Área temática: indicar el área temática en la que se inscribe el contenido de la comunicación o ponencia.

1. INTRODUCCIÓN

Una de las necesidades básicas de cualquier organización es conocer el estado de sus sistemas de información y decidir cuál es el nivel de gestión y control necesario para que las Tecnologías de la Información (TI) aporten realmente valor a la organización. La obtención de una visión objetiva del nivel de desempeño es una clara necesidad, se requiere una medición del estado y la identificación de áreas donde se deben realizar mejoras, implementando además un conjunto de herramientas de gestión para monitorizar esas mejorías. Los gestores de organizaciones, tanto públicas como privadas, se plantean con frecuencia la realización de análisis sobre el nivel de gestión de los procesos de TI, y desarrollan planes de negocio que buscan lograr un nivel adecuado de gestión y control sobre las infraestructuras de la información.

La evaluación de las capacidades y los procesos tecnológicos puede realizarse mediante modelos de madurez haciendo uso de herramientas como COBIT (ISACA, 2012), que define un modelo de gobernanza de TI elaborado por ISACA (*Information Systems Audit and Control Association*). En este trabajo se hace uso precisamente de COBIT para identificar procesos clave de gestión de las TI y analizar sus niveles de madurez en una muestra de 78 empresas brasileñas. Posteriormente, mediante técnicas de minería de datos se ha tratado de analizar y descubrir reglas que permitan conocer mejor el nivel de madurez tecnológica en las empresas de la muestra.

Control Objectives for Information and Related Technology (COBIT) es un modelo de gobernanza de TI que ayuda a que las TI aporten valor a las organizaciones y a identificar y gestionar adecuadamente los riesgos asociados al uso de la tecnología. COBIT es una guía de buenas prácticas dirigida al control y supervisión de las TI en las organizaciones (ITGI, 2007) que incluye además un metamodelo ontológico para la gobernanza de TI (Neto & Neto, 2013).

COBIT identifica una serie de procesos clave e incorpora un modelo de madurez basado en una escala de valoración discreta. De esta forma cada proceso recibe una asignación de puntos entre el nivel 0 (inexistente) y el nivel 5 (optimizado). Este enfoque se basa en el modelo de madurez definido por el *Software Engineering Institute* para la valoración de la capacidad de desarrollo de software.



Figura 1: Niveles de madurez de COBIT.

Con carácter general, este modelo de madurez considera una valoración “inexistente” (0) cuando hay una falta total de un proceso reconocible, y la organización ni siquiera ha identificado la existencia de un problema a resolver. Cuando ese problema ha sido identificado, pero no hay procesos estandarizados sino únicamente métodos aplicados de forma individual o caso por caso, se considera un nivel de madurez “inicial” (1). En el nivel “repetitivo” (2) los procesos se han desarrollado hasta el punto en que diferentes personas siguen procedimientos similares emprendiendo la misma tarea; sin embargo, no hay capacitación o comunicación formal de procedimientos estándar y la responsabilidad se deja a la persona. Se considera un proceso en estado “definido” (3) cuando se encuentra estandarizado y documentado, aunque los procedimientos no son sofisticados sino la formalización de las prácticas existentes. Cuando es posible monitorear y medir el cumplimiento de los procedimientos, emprender acciones correctoras si es preciso, y los procesos están bajo constante mejoramiento, se considera un nivel de madurez “gerenciado” (4). Finalmente, en el nivel “optimizado” (5) los procesos han sido refinados hasta un nivel de la mejor práctica, basados en los resultados de mejoramiento continuo y diseño de la madurez con otras organizaciones, usando las TI se usa de forma integrada para automatizar el flujo de trabajo. Mediante esta escala de madurez los responsables de la empresa pueden identificar dónde existen deficiencias en la administración de las TI dentro de la organización, además de obtener pautas y guías de buenas prácticas para mejorar los correspondientes procesos. El marco de trabajo COBIT aporta pautas para otorgar la correspondiente valoración a cada uno de los procesos que en él se identifican.

ISACA publicó la actual versión, COBIT 5, in 2012, sin embargo en este trabajo se optó por utilizar una versión previa, la versión 4.1, debido a la posibilidad de utilizar datos de una muestra de empresas brasileñas con las que se ha colaborado en un proyecto previo. En (De Haes et al, 2013) pueden encontrarse algunas cuestiones clave sobre la previsible evolución futura en gobernanza de TI y COBIT 5.0. Por otro lado, (Preittigun et al, 2012) realizan un análisis comparativo de la actual versión de COBIT con respecto a la versión previa 4.1. Esta versión previa define un modelo de procesos dividiendo la gestión de TI en cuatro grandes dominios con un total de 34 procesos. Los cuatro dominios son: “planear y organizar”, “adquirir e implementar”, “entregar y dar soporte”, y finalmente “monitorear y controlar”. Dentro de cada dominio se incluyen una serie de procesos básicos junto con pautas para asignar a cada uno de ellos un nivel de madurez en función del desarrollo que tiene en cada organización. La lista completa de procesos definidos en COBIT 4.1 se incluye en el anexo de este trabajo.

A partir del análisis de los niveles de madurez de cada proceso, la organización puede establecer sus metas y objetivos para mejorar los procesos relacionados con la gestión de las TI. Para Rafeq (2010), una medición de los niveles de madurez es necesaria para identificar carencias en procesos específicos, evaluar el estado actual del uso de las TI dentro de la organización y desarrollar planes de actuación orientados al logro de los objetivos planteados. Por otro lado, las técnicas de minería de datos pueden permitir identificar relaciones entre procesos, seleccionar procesos clave que puedan ser significativos a la hora de valorar globalmente el modelo de gobernanza de TI de cada organización, y poder realizar análisis comparativos sobre situaciones en otras empresas del sector o de características similares.

2. GENERACIÓN DE NUEVO CONOCIMIENTO CON MINERÍA DE DATOS

En la actualidad existen diferentes procesos o estrategias que permiten extraer y crear nuevo conocimiento mediante el análisis de la información que se genera en las organizaciones o procede de fuentes externas. Conscientes de ello, las empresas realizan grandes inversiones en este aspecto, considerando de vital importancia la obtención de nuevo conocimiento que pueda ser integrado y utilizado en los procesos de toma de decisiones (Tan et al, 2009). La minería de datos o data mining (DM) es una de esas

estrategias y comprende un conjunto de algoritmos y métodos que permiten explorar y analizar de forma automática o semi-automática grandes volúmenes de información (Witten & Frank, 2005). Estas técnicas pueden combinarse con técnicas estadísticas más tradicionales e incluso con metodologías propias de la Inteligencia Artificial para mejorar los procesos de toma de decisiones.

Las técnicas de minería de datos se incluyen dentro de un campo científico conocido como Descubrimiento de Conocimiento en Bases de Datos (*Knowledge Discovery in Database*, KDD), cuyo objetivo es la extracción de información y conocimientos útiles mediante técnicas de análisis inteligente y automatizado de conjuntos de datos relevantes, ofreciendo una mejor visión del negocio o de la propia organización (Seng & Chen, 2010). En (Ferri *et al*, 2004) puede encontrarse una introducción a las principales técnicas de minería de datos y sus campos de aplicación. Dentro de la MD se incluyen algoritmos y métodos que han demostrado ser muy útiles en la obtención de nuevo conocimiento (Witten & Frank, 2005); posibilitan la comprensión de comportamientos, generan pronósticos, identifican tendencias o cambios en esas tendencias, detectan patrones de comportamiento y también permiten descubrir relaciones entre la información para optimizar la toma de decisiones en las organizaciones, y en definitiva, para lograr alcanzar posiciones de ventaja competitiva. Un adecuado análisis de los datos disponibles mediante técnicas de MD puede proporcionar a las organizaciones importantes ventajas: incremento de ganancias, maximización de la eficiencia operativa, reducción de costes, generación de nuevas oportunidades de negocio, estimación del potencial económico de los clientes, mejora de la satisfacción de los mismos, descubrimiento de nuevos segmentos de clientes potenciales o nichos de mercado, etc. Los métodos de MD contribuyen a mejorar el control de la empresa, reduciendo el riesgo de predecir situaciones no deseadas y proporcionando recomendaciones basadas en experiencias anteriores (Bajo *et al.*, 2012). De acuerdo a Shmueli y Koppius (2011), el poder predictivo se refiere a la capacidad de un modelo para generar predicciones de observaciones en un período futuro u observaciones que no se incluyeron en la muestra original.

Un proyecto de MD debe completar una serie de etapas básicas; la primera consiste en determinar los objetivos que se quieren alcanzar y en obtener los datos a procesar. La etapa de pre-procesamiento consistirá en limpiar, transformar y reducir los

datos. Las operaciones que transforman los datos pueden ser la reducción de la dimensión, la discretización (conversión de un valor numérico a un valor nominal ordenado) y normalización (transformación de los valores de las variables cuando tienden a tener rangos distintos). Esta etapa es la que consume más de la mitad del tiempo del esfuerzo de desarrollo. La siguiente etapa consiste en aplicar la técnica de explotación de datos, determinando el modelo a utilizar, realizando análisis estadísticos y visualizando gráficamente los datos para tener una primera aproximación. Finalmente, la última etapa es la de análisis y evaluación de resultados obtenidos, verifica si los resultados son coherentes y los coteja con lo analizado estadística y gráficamente.

Como ya se ha comentado, a la hora de aplicar técnicas de MD hay una gran variedad de algoritmos disponibles. En concreto en este trabajo se han utilizado específicamente tres tipos de algoritmos para analizar datos sobre niveles de madurez tecnológica en una muestra de 78 empresas brasileñas:

- Clustering: técnicas que buscan clasificar conjuntos de instancias en grupos homogéneos. Los objetos se agrupan basándose en el principio de maximizar la similitud *intra-cluster* y minimizar la similitud *inter-clusters*. En este caso se ha aplicado a la identificación de grupos de empresas con modelos de gobernanza de TI similares.
- Selección de atributos: técnicas que permiten reducir la dimensión del espacio de trabajo, seleccionando un subconjunto de rasgos o atributos que representen correctamente a las diferentes instancias o elementos de la muestra. En este caso se ha buscado identificar aquellos procesos de COBIT que puedan resultar más significativos o determinantes a la hora de realizar un diagnóstico o valoración del modelo de gobernanza de TI de las organizaciones.
- Reglas de asociación: se trata de buscar los hechos que ocurren en común dentro de un conjunto de datos. Se identifican las correlaciones o interdependencias que existen entre los objetos. Como caso particular se podría plantear el análisis de interdependencias entre niveles de madurez de los procesos definidos por COBIT.

En definitiva, con la utilización de las técnicas enumeradas se proporcionará a los gerentes o directivos información que les permita valorar comparativamente los niveles de madurez tecnológica y desarrollar nuevas estrategias para mejorar sus procesos de TI.

3. METODOLOGÍA

En este trabajo se sigue el enfoque metodológico *Design Science* (DS) (Simon, 1996; Romme, 2007; Hevner & March, 2004), con dos fases diferenciadas: recopilación y estructuración de datos, y análisis de los mismos. Para recopilar los datos se diseñó un cuestionario dirigido a responsables y empleados de los departamentos de TI de empresas de diferentes sectores. El cuestionario fue implementado sobre la herramienta *Google Drive*, y se encontraba estructurado en dos partes: la primera, más descriptiva, sobre la organización y perfil del encuestado, y la segunda parte en la que se solicitaba a los encuestados que realizaran una valoración de los niveles de madurez en su organización de los 34 procesos definidos en COBIT 4.1. Para esa valoración se les aportaban pautas e indicaciones, con una tabla explicativa de cada nivel de madurez según el modelo del *Software Engineering Institute*, contándose igualmente con la colaboración de estudiantes de postgrado de la Universidad do Vale do Rio dos Sinos (UNISINOS) de Brasil

Los cuestionarios fueron respondidos a través de la Web por medio de la herramienta de generación de encuestas de *Google Drive*. Además de la tabulación de los diferentes niveles de madurez de los 34 procesos, se obtuvieron los nombres de las empresas, sectores en los que operan, las ciudades sede, los cargos y departamentos de las personas encuestadas, y las fechas de realización de las encuestas.

En total se recopilaron datos de ochenta y cuatro (84) empresas que operan en el estado de Rio Grande do Sul, en Brasil; de ellas, seis (6) enviaron respuestas incompletas y fueron eliminadas, resultando una muestra final de setenta y ocho (78) empresas válidas. Las respuestas válidas fueron organizadas a través de una tabla bidimensional, y se generó un archivo ARFF para su tratamiento con el software *WEKA - Waikato Environment for Knowledge*. En la muestra analizada predominaron empresas pertenecientes a los sectores de servicios, venta al por menor e industrias

metalmecánicas. En cuanto al tamaño de las empresas de la muestra, la mayoría son clasificadas como grandes o de tamaño medio. Debe tenerse en cuenta que se buscó empresas en las que existiera un departamento de TI para poder realizar el análisis de los procesos definidos por COBIT.

Una vez recopilados y procesados los datos, en la etapa de análisis se utilizan diversas técnicas de minería de datos: por un lado el algoritmo de agrupamiento *EM*, además de diferentes métodos de selección de atributos y el algoritmo *Apriori* para la generación de reglas de asociación entre procesos. Los análisis realizados se dividieron en tres partes: en la primera parte se utilizó la técnica no supervisada de clustering algoritmo EM (*Expectation Maximization*), tratando de buscar clusters (agrupamientos) entre los datos, es decir, grupos de empresas como modelos de gobernanza de TI similares. En (Rocha et al, 2013) también se utilizaron con éxito estas mismas técnicas de agrupamiento para la identificación de metáforas organizacionales en una muestra de empresas. El algoritmo EM es un método probabilístico con una técnica iterativa que maximiza la esperanza de la función de verosimilitud de un determinado vector de parámetros sobre una muestra en la que hay ciertos datos ocultos (Sierra, 2006). Una de las ventajas de este método es la identificación automática del número más apropiado de clusters. El algoritmo EM establece en cada instancia una distribución de probabilidad de pertenencia a cada cluster, de manera que se pueda decidir el número de clusters a generar a partir de una validación cruzada. Utiliza el modelo finito combinado de Gauss, suponiendo que todos los atributos son variables aleatorias independientes.

En una segunda parte del análisis se utilizaron diversas técnicas de selección de atributos que se encuentran implementadas en el software *Weka*. Estas técnicas tratan de identificar los subconjuntos de atributos que mejor puedan clasificar la clase de cada instancia. Esta selección de atributos tiene dos componentes principales: un método de evaluación, que determina la calidad del conjunto de atributos para discriminar la clase, y un método de búsqueda, que determina la forma de realizar la búsqueda de conjuntos. Se pueden distinguir dos categorías de métodos de evaluación, en la primera se utiliza directamente un clasificador específico para medir la calidad del subconjunto de atributos a través de la tasa de error del clasificador. Estos métodos necesitan un proceso completo de entrenamiento y evaluación en cada caso de búsqueda, por eso resultan de un elevado coste computacional. La alternativa es la utilización de métodos que no utilizan un clasificador específico, por ejemplo el método *CfsSubsetEval* que se

encuentra implementado en *Weka* y que se basa en calcular la correlación de la clase con cada atributo, y eliminar atributos que tienen una correlación muy alta con atributos redundantes. Según este método los subconjuntos preferidos son aquellos altamente correlacionados con el atributo que define las clases y con poca correlación entre ellos. El método de búsqueda determina la forma de realizar la búsqueda de conjuntos. La evaluación exhaustiva de todos los posibles subconjuntos se convierte en un problema combinatorio inabordable cuando el número de atributos es elevado. Por tanto, se necesitan estrategias de búsqueda más eficientes. Una de las estrategias más efectiva, por su rapidez, es el *ForwardSelection*, que se basa en elegir primero el mejor atributo, y realizar un proceso iterativo de ir añadiendo atributos que aporten más información hasta llegar a la situación en la que añadir un nuevo atributo empeora la situación.

En la tercera parte del análisis se buscó la identificación de reglas de asociación. Los algoritmos de asociación permiten la búsqueda automática de reglas que relacionen conjuntos de atributos entre sí. Son algoritmos no supervisados, es decir que no existen relaciones conocidas a priori con las que contrastar la validez de los resultados, sino que se evalúa si esas reglas son estadísticamente significativas. *Weka* tiene implementado, entre otros, el algoritmo *Apriori* para obtener reglas de asociación. Para evaluar la efectividad del algoritmo se utilizan dos medidas de calidad: el soporte, también conocido como cobertura, que indica el número de instancias que la regla predice correctamente; y la confianza que indica el porcentaje de veces que la regla se cumple. Se puede configurar el algoritmo con varias opciones: la opción *UpperBoundSupport* indica el límite superior de cobertura requerido para aceptar un conjunto de reglas. Si no se encuentran conjuntos de instancias suficientes que cumplan las reglas requeridas se va disminuyendo el límite hasta llegar al límite inferior (opción *LowerBoundSupport*). La opción *minmetric* indica la confianza mínima para mostrar una regla y la opción *numRules* indica el número máximo de reglas que se desea que se generen.

4. APLICACIÓN DE LAS TÉCNICAS DE MINERÍA DE DATOS

4.1. Análisis de clusters

Para realizar el análisis de clusters se consideraron los 34 procesos de COBIT, considerando la escala de valoración de tipo numérico en el rango 0-5, tal como se

establece en los niveles de madurez. Tras aplicar el algoritmo de clustering EM, los resultados obtenidos son los que se resumen en la Tabla 1. En dicha tabla los procesos quedan identificados con un código alfanumérico de acuerdo a la lista que se incluye en el Anexo. Como puede apreciarse, el algoritmo identificó automáticamente 4 grupos de empresas. En la Tabla 1 cada grupo se representa a partir de los valores medios de los 34 procesos en las empresas del grupo.

A nivel de procesos, pueden observarse algunos procesos que alcanzan buenos niveles de madurez en todos los grupos. Dentro de ellos estarían, por ejemplo, los procesos relativos a la garantía de seguridad de los sistemas (DS5), la administración de servicios de terceros (DS2), o la garantía de continuidad de los servicios (DS4). Por el contrario, el proceso relativo a la administración del personal de TI (PO7) tiene promedios muy bajos en todos los grupos, incluyendo 3 de ellos como los más bajos (0, 1 y 2). Los procesos relacionados con la comunicación de los objetivos y directrices de gestión (PO6) y el entrenamiento de usuarios (DS7) también están entre los procesos con niveles de madurez más bajos. Este primer análisis de resultados viene a confirmar la preocupación de las empresas en general por los aspectos relacionados con la seguridad de los sistemas de información y la garantía de continuidad de servicio. Quizás se observa como las empresas dan prioridad a aspectos relacionados con la tecnología frente a aspectos más relacionados con la gestión de los recursos humanos, pudiéndose generar vulnerabilidades y riesgos precisamente por no realizar una adecuada gestión de estos últimos.

En la Tabla 1 pueden observarse los grupos identificados de acuerdo a los niveles de madurez de los procesos de COBIT. La distribución porcentual de las empresas en los grupos está relativamente equilibrada, quedando el cluster 1 con un tamaño ligeramente mayor (35% de las empresas consideradas) y el cluster 3 ligeramente más pequeño (17%). El cluster 3 corresponde precisamente a las empresas como menores niveles de madurez tecnológica, mientras que el grupo con mejores niveles corresponde al cluster 0, en el que quedan encuadradas el 24% de las empresas analizadas. La Tabla 2 ofrece más información sobre las características de los grupos identificados, indicando el número de empresas en cada grupo e información estadística sobre los niveles de madurez de los procesos, además de una primera caracterización sobre las empresas de cada grupo desde el punto de vista de la gobernanza de TI.

	<i>Cluster 0</i> 24%	<i>Cluster 1</i> 35%	<i>Cluster 2</i> 24%	<i>Cluster 3</i> 17%
<i>Código del proceso</i>	<i>Valores medios para cada proceso</i>			
PO1	3,7369	1,80960	2,68620	0,61540
PO2	3,7895	1,33280	2,84320	0,61540
PO3	4,3157	1,66520	3,25550	0,76920
PO4	3,9999	1,29320	2,69100	0,76920
PO5	4,3685	2,44640	3,46450	0,76920
PO6	3,7895	1,32600	2,68620	0,00000
PO7	3,6317	0,73070	1,43100	0,07690
PO8	4,0527	1,44200	2,99380	1,07690
PO9	4,2633	1,40920	3,03990	0,53850
PO10	4,2632	1,44200	3,20310	0,46150
AI1	4,3684	1,70240	3,51700	1,46150
AI2	3,6843	2,29320	3,47070	0,69230
AI3	4,0525	1,73960	3,20320	0,76920
AI4	3,9473	1,33040	2,05260	0,30770
AI5	4,2631	2,10720	3,10480	0,69230
AI6	3,842	1,92560	3,15090	0,38460
AI7	4,2105	1,59530	2,77840	0,38460
DS1	4	1,95840	3,00010	0,38460
DS2	4,5264	2,40480	3,57540	0,92310
DS3	4,0002	1,73960	3,25520	0,38460
DS4	4,5838	2,66960	3,56920	1,07690
DS5	4,6315	2,92560	3,83700	0,84620
DS6	4,0527	2,10720	2,89540	0,46150
DS7	3,7895	2,03720	2,00020	0,23080
DS8	4,421	1,51200	2,84330	0,46150
DS9	4,2106	1,55370	3,15070	0,53850
DS10	4,421	1,88840	2,68020	0,30770
DS11	4,2631	1,89280	3,06800	0,53850
DS12	3,9475	1,80960	2,84310	0,38460
DS13	4,3685	2,03280	3,20930	0,53850
MO1	4,3686	1,58640	2,63380	0,07690
MO2	4,3685	1,44200	2,88920	0,69230
MO3	4,3685	2,24730	3,41850	0,53850
MO4	4,2107	1,58640	3,05220	0,46150

Tabla 1: Resultados del agrupamiento con el algoritmo EM. Valores medios de los niveles de madurez para cada uno de los 34 procesos.

Cluster	Número de empresas	Niveles de madurez medios	Niveles de madurez mínimos	Niveles de madurez máximos	Características de las empresas
0	19	4,1503	3,6317	4,6315	Tienen sus procesos definidos, bien administrados y optimizados
1	27	1,7936	0,7307	2,9256	Los procesos no están bien gestionados, niveles iniciales, procesos repetidos y poco definidos
2	19	2,9851	1,431	3,837	Los procesos no están bien gestionados, procesos repetidos y solo algunos procesos están definidos
3	13	0,5656	0	1,4615	Los procesos no están bien gestionados, inexistentes o en niveles iniciales

Tabla 2: Descripción de las principales características de los clusters.

También puede observarse como más de la mitad de las empresas analizadas (las que fueron asignadas a los clusters 1 y 3) poseen niveles de madurez tecnológica que varían entre inexistentes y repetitivos, es decir, los procesos están pobremente gestionados y en niveles muy iniciales. Realmente solo en la cuarta parte de las empresas, las correspondientes al cluster 0, los procesos están correctamente gestionados. Analizando las características de este grupo, se pudo observar cómo la mayoría de las empresas correspondían a las de tamaño grande, con departamentos de TI mejor dotados y más estructurados. Otra característica de las empresas de este grupo es la menor variabilidad en los niveles de madurez entre los procesos. En otros grupos se observa una mayor variabilidad, lo que pone de manifiesto una mayor preocupación por determinados procesos en detrimento de otros.

Las Figuras 2 y 3 muestran una serie de histogramas generados por *Weka* con los niveles de madurez en los 34 procesos de COBIT y diferenciando el cluster al que fue asignado cada empresa. Los colores en los gráficos identifican a cada uno de los cuatro grupos creados por el algoritmo de clustering.



Figura 2: Distribución de las empresas en los clusters.

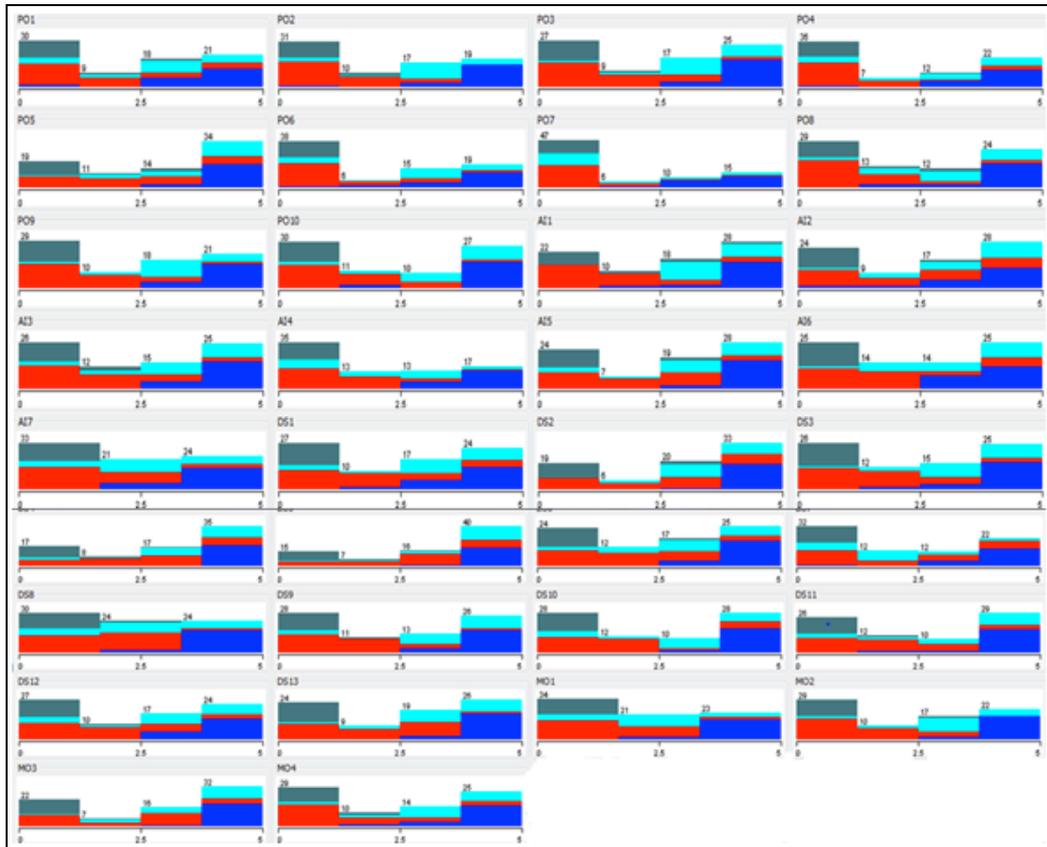


Figura 3: Histogramas de los procesos en cada cluster.

A la vista de los gráficos de la Figura 3, se podría afirmar que los niveles de madurez de los procesos se encuentran habitualmente en niveles altos o bajos, con una menor distribución en los valores intermedios. Además se observa como en las empresas del cluster 2 hay una mayor variabilidad en esos niveles de madurez en los diferentes procesos.

4.2. Selección de atributos

Tras el análisis de clustering, el siguiente objetivo que se planteó es tratar de seleccionar un subconjunto de procesos que representen de manera fiel la estructura de grupos presente en la muestra. De los 34 procesos no todos son igualmente válidos para predecir el nivel de madurez tecnológica de una organización, de manera que se planteó seleccionar aquellos con mayor poder discriminante.

Weka incorpora una gran variedad de métodos de selección de atributos, en la Tabla 3 se muestran los subconjuntos seleccionados usando el método de evaluación *CfsSubsetEval* y diferentes métodos de búsqueda. En todos los casos el atributo que define la clase es el que asigna cada empresa al cluster de acuerdo al análisis anterior. Como puede observarse, aunque los subconjuntos obtenidos no son siempre iguales, sí que tienen grandes semejanzas, de manera que en la última fila de la tabla se muestran los 21 atributos que fueron seleccionados por todos los métodos utilizados.

Como puede comprobarse, a la hora de analizar el nivel de madurez tecnológica de las empresas, al menos en la muestra analizada, únicamente sería necesario estudiar el nivel de madurez de 21 de los 34 procesos de COBIT. Además se observa cómo se han seleccionado procesos de cada uno de los 4 dominios principales de COBIT. De esta manera, Aquellos procesos que han sido eliminados por los métodos de selección de atributos están fuertemente correlacionados con los procesos seleccionados, y, por tanto, desde el punto de vista de la clasificación de las empresas tampoco aportan mucha información.

Método de búsqueda	Nº atributos	Atributos
Best first	23	PO2, PO3, PO5, PO6, PO7, PO8, PO9, PO10, AI1, AI4, AI6, DS1, DS2, DS3, DS6, DS7, DS8, DS9, DS10, DS13, MO1, MO2, MO3
SubsetSizeForwardSelection	23	PO2, PO3, PO5, PO6, PO7, PO8, PO9, PO10, AI1, AI4, AI6, DS1, DS2, DS3, DS6, DS7, DS8, DS9, DS10, DS13, MO1, MO2, MO3
GeneticSearch	24	PO2, PO3, PO5, PO6, PO7, PO8, PO9, PO10, AI1, AI4, AI5, AI6, DS1, DS2, DS4, DS6, DS7, DS8, DS9, DS10, DS13, MO1, MO2, MO3
GreedyStepwise	23	PO2, PO3, PO5, PO6, PO7, PO8, PO9, PO10, AI1, AI4, AI6, DS1, DS2, DS3, DS6, DS7, DS8, DS9, DS10, DS13,

		MO1, MO2, MO3
LinearForwardSelection	23	PO2, PO3, PO5, PO6, PO7, PO8, PO9, PO10, AI1, AI4, AI6, DS1, DS2, DS3, DS6, DS7, DS8, DS9, DS10, DS13, MO1, MO2, MO3
RankSearch	34	PO1, PO2, PO3, PO4, PO5, PO6, PO7, PO8, PO9, PO10, AI1, AI2, AI3, AI4, AI5, AI6, AI7, DS1, DS2, DS3, DS4, DS5, DS6, DS7, DS8, DS9, DS10, DS11, DS12, DS13, MO1, MO2, MO3, MO4
ScatterSearchV1	23	PO2, PO3, PO5, PO6, PO7, PO8, PO9, PO10, AI1, AI4, AI6, DS1, DS2, DS3, DS6, DS7, DS8, DS9, DS10, DS13, MO1, MO2, MO3
Atributos seleccionados por todos los métodos	21	PO2, PO3, PO5, PO6, PO7, PO8, PO9, PO10, AI1, AI4, DS1, DS2, DS6, DS7, DS8, DS9, DS10, DS13, MO1, MO2, MO3

Tabla 3: Selección de los procesos más significativos a la hora de valorar la madurez tecnológica de las empresas.

4.3. Búsqueda de reglas de asociación

Finalmente, sobre el conjunto de atributos seleccionados se trató de buscar asociaciones entre los niveles de madurez mediante el algoritmo *Apriori*. El objetivo era analizar si el nivel de algún proceso tenía influencia sobre los niveles de otros procesos. Para ello se optó por categorizar los valores de madurez en tres niveles: bajo, medio y alto, considerando nivel bajo a los procesos en estados inexistentes o iniciales (valores de madurez 0 y 1), medio a los procesos repetitivos o definidos (valores 2 y 3), y alto a los procesos administrados u optimizados (valores 4 y 5). Para la aplicación del algoritmo de generación de reglas de asociación fueron establecidas las siguientes condiciones:

- a) Soporte mínimo del 25% de registros, para considerar una regla válida era necesario que se observara en una cuarta parte de las 78 empresas.
- b) Confianza mínima de 0,9, admitiéndose por tanto un error del 10% en la verificación de la regla.

Con esas condiciones el algoritmo generó un total de 306 reglas, de las cuales fueron eliminadas las redundantes y aquellas que, por el elevado número de atributos

que intervenían, resultaban complejas de interpretar. Las reglas finalmente seleccionadas son las que se muestran en la Tabla 4.

<i>Nº regla</i>	<i>Si...</i>	<i>Entonces</i>	<i>Soporte</i>	<i>Confianza</i>
1	PO10=Bajo	PO7=Bajo	30	1,00
2	AI4=Bajo	PO7=Bajo	34	0,97
3	DS13=Bajo	MO1=Bajo	23	0,96
4	MO2=Alto	MO3=Alto	21	0,95
5	PO6=Bajo	PO7=Bajo	36	0,95
6	MO1=Bajo	PO7=Bajo	32	0,94
7	PO2=Bajo	PO7=Bajo	29	0,94
8	DS8=Bajo	PO7=Bajo	28	0,93
9	PO8=Bajo	PO7=Bajo	27	0,93
10	MO2=Bajo	PO7=Bajo	27	0,93
11	PO3=Bajo	PO7=Bajo	25	0,93
12	DS6=Bajo	PO7=Bajo	22	0,92
13	DS13=Bajo	PO7=Bajo	22	0,92
14	DS13=Bajo	PO7=Bajo MO1=Bajo	22	0,92
15	MO1=Alto	PO5=Alto	21	0,91
16	MO1=Alto	DS2=Alto	21	0,91
17	AI1=Bajo	PO7=Bajo	20	0,91
18	MO3=Bajo	PO7=Bajo	20	0,91
19	MO3=Bajo	MO2=Bajo	20	0,91
20	PO10=Bajo	PO6=Bajo	27	0,90

Tabla 4: Reglas de asociación entre niveles de los procesos.

Como puede verse, la mayor parte de las reglas obtenidas (17 de 20) están vinculadas con niveles de valoración baja en los procesos, y únicamente en 3 de ellas intervienen procesos de alta valoración. Es importante puntualizar que las reglas se deben aplicar únicamente en el sentido indicado (*si ... entonces*). Muchas de las reglas permiten deducir niveles de madurez bajos en el proceso relativo a la administración de los recursos humanos de TI (proceso PO7) cuando los siguientes procesos se encuentran también en estados iniciales:

- Procesos del dominio PO: PO2, PO3, PO6, PO8, PO10;
- Procesos del dominio DS: DS6, DS8, DS13;
- Procesos del dominio AI: AI1 AI4;
- Procesos del dominio MO: MO1, MO2, MO3.

La vinculación del proceso PO7 con otros procesos del mismo dominio resultan más evidentes, así por ejemplo cuando existe un nivel inadecuado o bajo de madurez en la gestión de proyectos de TI (PO10) parece obvio que la gestión de los recursos humanos (PO7) también tenga unos niveles de madurez bajos.

5. CONCLUSIONES FINALES

La utilización de técnicas de minería de datos sobre los resultados de la medición de niveles de madurez de procesos tecnológicos en diferentes empresas, permite a los responsables de TI obtener una visión comparativa sobre el modelo de gobernanza de TI que están aplicando en sus organizaciones. Gracias a estos análisis se pueden identificar aquellos procesos que necesitan optimizarse, o reconocer relaciones entre los procesos y dependencias entre los respectivos niveles de madurez. Puede ser también importante la generación de predicciones de niveles de desarrollo en nuevos procesos, con el objeto de desarrollar estrategias de mejora de procesos organizacionales y su utilización en la toma de decisiones.

En este estudio, gracias a estas técnicas, se ha podido identificar 4 grupos o patrones de gobernanza de TI en una muestra de empresas brasileñas que operan en el estado de Rio Grande do Sul. Además se ha podido comprobar, que para hacerse una idea global de esos modelos de gobernanza, se puede seleccionar un conjunto limitado de procesos, de entre los que se definen en el marco de trabajo COBIT. A su vez, sobre ese conjunto se han podido obtener una serie de reglas de asociación que aportan información para comprender la influencia de unas variables sobre otras, observándose como la mayor parte de las reglas están vinculadas a relaciones entre procesos que alcanzan niveles de madurez bajos.

Toda esta información, convenientemente utilizada y contextualizada en cada organización, permite a las organizaciones un mejor conocimiento sobre su funcionamiento interno, identificar posibles carencias, mejorar la transparencia interna, minimizar los riesgos y direccionar mejor sus inversiones tecnológicas.

El estudio realizado sobre una muestra de empresas brasileñas podría ampliarse, considerando una muestra mayor de empresas, realizando análisis sectoriales o análisis comparativos con otras regiones o países. Igualmente sería interesante disponer de una

muestra mayor para poder analizar la influencia del factor tamaño o sector de operación en los niveles de madurez de procesos tecnológicos. Todo ello se plantea como trabajo futuro de colaboración entre las dos instituciones a las que pertenecen los autores de este trabajo.

6. REFERENCIAS BIBLIOGRÁFICAS

- BAJO, J.; BORRAJO, M.L.; PAZ, J.F.D.; CORCHADO, J.M. y PELLICER, M.A. (2012). “A multi-agent system for web-based risk management in small and medium business”. *Expert Systems with Applications*, v. 39, n. 8, pp 6921–6931.
- DE HAES, S.; VAN GREMBERGEN, W. y DEBRECENY, R. (2013). “COBIT 5 and Enterprise Governance of Information Technology: Building Blocks and Research Opportunities”. *Journal of Information Systems*, 27(1), 307.
- FERRI, C.; RAMÍREZ, M. J.; HERNÁNDEZ, J. (2004). “Introducción a la minería de datos”. Prentice Hall.
- HEVNER, A. y MARCH, S. (2004). “Design Science in information systems research”. *MIS Quarterly*. Vol.28, no.1, pp. 75-105.
- ISACA (2012). Information Systems Audit and Control Association. COBIT Five: A Business Framework for the Governance and Management of Enterprise IT. Rolling Meadows, IL: ISACA.
- ITGI (2012). Information Technology Governance Institute. CobiT 4.1 modelo, objetivos de control, directrices de gestión y modelo de madurez. Disponible en: <<http://www.isaca.org/Knowledge-Center/>>.
- NETO, J.S. y NETO, A.N.F. (2013). “Metamodel of the IT governance framework COBIT”. *Journal of Information Systems and Technology Management*, 10(3), 521-540.
- PREITTIGUN, A.; CHANTATUB, W. y VATANASAKDAKUL, S. (2012). “A Comparison between IT Governance Research and Concepts in COBIT 5”. *International Journal of Research in Management & Technology*, 2(6), 581-590.

- RAFEQ, A. (2010). “Using COBIT for Assessing IT Process Maturity: A Case Study”. COBIT Focus, v. 4.
- ROCHA, R.; COBO, A.; JOHANN, S. y VANTI, A. (2013). “Imágenes organizacionales de las empresas brasileñas: detección y análisis con técnicas de minería de datos”. Revista de Ciências da Administração, v. 15, p. 105-120.
- ROMME, A.G.L. y DAMEN, I.C.M. (2007). “Toward Science-Based Design in Organization Development: Codifying the Process”. The Journal of Applied Behavioral Science, v. 43, n. 1, p. 108-121.
- SENG, J.L. y CHEN, T.C. (2010). “An analytic approach to select data mining for business decision”. Expert Systems with Applications, v.37, n.12, p.8042-8057.
- SHMUELI, G. y KOPPIUS, O.R. (2011). “Predictive analytics in information systems research”. MIS Quarterly, v.35, n.3.
- SIERRA, B. (2006). “Aprendizaje Automático: Conceptos básicos y avanzados”. Pearson Prentice Hall.
- SIMON, H.A. (1996). The Sciences of the Artificial. USA: MIT Press.
- TAN, P; STEINBACH, M.; y KUMAR, V. (2009). Introduction to data mining. USA: Addison-Wesley Longman Publishing.
- WITTEN, I y FRANK, E. (2005) Data mining: practical machine learning tools and techniques. New Zealand: Morgan Kaufmann Publishers.

ANEXO: PROCESOS DEFINIDOS EN COBIT 4.1

PO1	Definir un plan estratégico de TI
PO2	Definir la arquitectura de la información
PO3	Determinar la dirección tecnológica
PO4	Definir los procesos, organización y relaciones de TI
PO5	Administrar la inversión en TI
PO6	Comunicar las aspiraciones y la dirección de la gerencia
PO7	Administrar los recursos humanos de TI
PO8	Gestión de la calidad
PO9	Evaluar y gestionar los riesgos de TI
PO10	Administrar proyectos
AI1	Identificar soluciones automatizadas
AI2	Adquirir y mantener software aplicativo
AI3	Adquirir y mantener infraestructura tecnológica
AI4	Facilitar la operación y el uso
AI5	Adquirir recursos de TI
AI6	Administrar cambios
AI7	Instalar y acreditar soluciones y cambios
DS1	Definir y administrar niveles de servicio
DS2	Administrar los servicios de terceros
DS3	Administrar el desempeño y la capacidad
DS4	Garantizar la continuidad del servicio
DS5	Garantizar la seguridad de los sistemas
DS6	Identificar y asignar costos
DS7	Educar y entrenar a los usuarios
DS8	Administrar la mesa de servicio y los incidentes
DS9	Administrar la configuración
DS10	Administrar los problemas
DS11	Gestionar los datos
DS12	Administrar el ambiente físico
DS13	Administrar las operaciones
ME1	Monitorear y evaluar el desempeño de TI
ME2	Monitorear y evaluar el control interno
ME3	Garantizar el cumplimiento regulatorio
ME4	Proporcionar gobierno de TI