

Optimización de Selección de Proveedores Integrando un Árbol de Decisión a un Proceso de Negocio¹

Providers Selection Optimization Integrating a Decision Tree in a Business Process

Andrés Paolo Castaño Vélez

*Doctorando(c) Ingeniería Informática
Especialista en Docencia Universitaria
Especialista en CAD/CAM
Ingeniero Electricista
Docente Universidad de Caldas
Grupo de Investigación GITIR
andres.castano@ucaldas.edu.co*

Carlos Alberto Ruiz Villa

*Msc en Bioingeniería
Especialista en CAD/CAM
Ingeniero Electricista
Docente Universidad de Caldas
Grupo de Investigación GITIR
Departamento de Informática y Computación.
Universidad Nacional de Colombia
carv@ucaldas.edu.co*

Andrés Mauricio Martínez

*Especialista en Telecomunicaciones
Ingeniero de Sistemas
Docente Universidad de Caldas
Grupo de Investigación GITIR
andres.martinez@ucaldas.edu.co*

Recibido Mayo 23 de 2011 – Aceptado Noviembre 30 de 2011

RESUMEN

La arquitectura orientada a servicios (SOA) permite abordar los nuevos retos empresariales, ser más competitivos y disponer de sistemas de

¹ Es un producto derivado del trabajo de investigación "Prototipo para la selección dinámica adaptativa de servicios proveedores en un ESB con minería de datos", realizado en el doctorado en Ingeniería Informática en la Universidad Pontificia de Salamanca en Madrid, bajo la Tutoría del Dr. Jesús Soto Carrión.

información integrados. Así mismo, tecnologías como la gestión de procesos de negocios (BPM), el monitoreo de las actividades de negocio (BAM) y los servicios web forman un complemento esencial para SOA.

En este trabajo se integran varias de estas tecnologías en una sola aplicación, las cuales permiten, en una etapa de un proceso de negocio, realizar el análisis de proveedores de insumos para la empresa a través de la generación de un árbol de decisión usando código embebido de la herramienta libre de minería de datos Weka, con el fin de retroalimentar el proceso de negocio para mejorarlo.

Como resultado se obtuvo un prototipo capaz de seleccionar proveedores a través de un árbol de decisión.

Palabras clave: Gestión de Procesos de Negocio (BPM), Minería de Datos, Bus de Servicios Empresariales (ESB), jBPM, Arquitectura Orientada a Servicios (SOA), servicio web.

ABSTRACT

Service-oriented architecture (SOA) is developed to address new business challenges; they have become more competitive and integrated information systems. In addition, technologies such as business process management (BPM), Business activity monitoring (BAM) and web services are an essential complement to SOA.

In this paper we integrate several of these technologies in a single application that allows, in a phase of a business process, to perform the analysis of input suppliers to the company, through the generation of a decision tree, using embedded code of free tool Weka data mining to feedback the business process to improve it.

The result was a prototype able to select suppliers through a decision tree.

Key words: Business Process Management (BPM), Data Mining, Enterprise Service Bus (ESB), jBPM, Service Oriented Architecture (SOA), web services.

1. INTRODUCCIÓN

Las tecnologías de la información, al igual que los diferentes aspectos de

la economía mundial, han evolucionado de tal forma en los últimos años, que ha llevado a las empresas a ser más competitivas y a centrar esfuerzos en la reducción de costos en todos sus frentes con el fin de maximizar las utilidades. Para ello se ha establecido una búsqueda constante en la optimización de los procesos de la empresa. Este campo supuso inicialmente una revolución en la década de los 90's por la forma de coordinar las actividades internas de las organizaciones, cambiando drásticamente y expandiéndose a entornos B2B (Business to Business), de forma que se facilitara la colaboración entre distintas organizaciones y hoy día se aplican en entornos diseñados de acuerdo al paradigma de la computación orientada al servicio (Krafzig, Banke y Slama 2004).

En este contexto aparecen los sistemas de gestión de procesos de negocio (BPMS) y los sistemas de gestión de workflow que gracias a su característica de reutilización permiten elaborar la lógica de un proceso de negocio adaptable a cualquier dominio. Aunque estas herramientas permiten tener prácticamente toda la lógica de negocio, existen pasos en estos procesos que aún siguen dependiendo del desempeño humano, en este punto surge la idea de hacer aportes a la automatización de un proceso de negocio que requiera la selección de proveedores de insumos, tarea que es realizada por un workflow humano y este resultado lo alimenta nuevamente al proceso de negocio para su continuación.

Aunque hasta ahora no se han diseñado prototipos que integren diversas tecnologías y permitan la asignación de proveedores, en este trabajo se analizan algunas publicaciones recientes y se abordan diferentes tecnologías para la realización e implementación de un prototipo, al mismo tiempo se expone la arquitectura propuesta para la selección de proveedores en un BPM a través de un árbol de decisión, las dificultades encontradas y finalmente se presentan las conclusiones y la posibilidad de desarrollar trabajos futuros.

2. MÉTODOS

Para mostrar cómo es posible desarrollar una aplicación que integre diferentes tecnologías y arroje resultados de forma transparente para el usuario, se diseñó una aplicación que permite, a partir de una suite BPM (jBPM), involucrar un proceso de minería de datos con una llamada realizada desde una etapa de un proceso de negocio y devolver los resultados a la etapa del proceso que hizo la llamada.

El BPMS contiene toda la lógica del proceso de negocio y hace uso de un árbol de decisión extraído de un aplicativo de minería de datos para realizar la selección; en todo el proceso se almacenaron los datos en una base que intercambia información con las diferentes tecnologías, usando para ello formatos de intercambio de datos estándares como XML o JSON.

Esta solución se apoya en la arquitectura orientada a servicios que permite que, aplicaciones desarrolladas con diferentes tecnologías, compartan datos y funcionalidades, facilitando la creación de nuevos sistemas con aplicaciones separadas que trabajan juntas de forma desacoplada.

a. Conceptualización y trabajos relacionados

Existen diferentes trabajos que se han enfocado en el tema de BPM y minería de datos, pero no como una aplicación integrada sino como procesos separados que proporcionan servicios. Se puede destacar el trabajo presentado por (Rinderle-Ma y van der Aalst 2007), en el que se introduce un enfoque de minería para la asignación de personal usando el registro de eventos, el conocimiento de las estructuras de la organización y un árbol de decisión como método de aprendizaje. El trabajo se centra en las reglas de minería para la asignación de personal, pero no consideran la optimización de la asignación de recursos con un entorno dinámico.

Por otra parte, (Lohrmann y Reichert 2010) hacen énfasis en la integración de la gestión de la calidad con conceptos de gestión de procesos desde una perspectiva empresarial. Ellos realizan una discusión sobre diferentes puntos de vista de calidad y su adecuación en la gestión de procesos de negocio, buscando establecer bases para la definición de procesos de negocio con calidad.

Hay una gran variedad de herramientas de minería de datos que los usuarios pueden usar para ayudarse en el procesamiento de datos, recuperación de información y procesos de generación de conocimiento, pero no es una tarea fácil para los usuarios escoger la mejor opción. Con este objetivo, (Seng y Chen 2010) desarrollan un modelo de selección para resolver este inconveniente, su modelo recomienda el algoritmo más adecuado para resolver un problema de negocio usando un procedimiento estándar. El modelo proporciona un algoritmo estándar como base de la modelización dinámica en minería de datos. Para llevarlo a cabo, tomaron un conjunto de 22 problemas comerciales referidos a la banca, finanzas, seguros, telecomunicaciones, comercio minorista y aplicaciones de manufactura y

los clasificaron en 12 categorías de aplicaciones de negocios de acuerdo a sus características.

Así mismo, (Ari, Li y Jain, y otros 2008), describen el diseño y la implementación de un sistema que gestiona un modelo de minería de datos de los activos de una organización que permite tomar decisiones y previsiones de los procesos de negocio en tiempo real. En otro trabajo del grupo (Ari, Li y Kozlov, y otros 2008) describen un modelo de gestión de minería de datos que aborda los problemas de caducidad del modelo, crecimiento y adaptación de la gestión, tiempo de comunicación entre las partes al presentarse cambios, vacíos semánticos en la interpretación de los modelos e integración de los procesos de negocio con apoyo sostenible y operacionalización.

El trabajo se basa conceptualmente en la arquitectura orientada a servicios (SOA) y su funcionamiento, a través de servicios como elementos fundamentales para integrar y desarrollar aplicaciones. Para este contexto se acogió la definición general desarrollada por (Krafzig, Banke y Slama 2004), según el cual, *“Una arquitectura de software como un conjunto de definiciones que describen los componentes de software y asignan la funcionalidad del sistema a estos componentes. Describe la estructura técnica, restricciones y características de los componentes y las interfaces entre ellos”*.

En el mismo sentido, otros autores, (Molina, y otros 2000) discuten el concepto de modelado de procesos de negocio y su importancia en el desarrollo de cualquier industria y como una mejora en un proceso de negocio afecta positivamente la producción y por consiguiente, los productos desarrollados.

Otros estudios hacen hincapié en las características que diferencian un BPM de un Workflow management system (WfMS), en la etapa de diagnóstico y definición de procesos, aspectos desatendidos en un WfMS, (Van der Aalst, Hofstede y Weske 2003).

La tendencia hoy es hacia un paradigma orientado a procesos, donde las aplicaciones deben cubrir la actividad global de la empresa y las herramientas son los sistemas de gestión de procesos de negocio (BPMS).

Para la integración de las diferentes aplicaciones se consideró un ESB, como lo define el estudio de (Desmet, y otros 2007) y se desarrolló un estándar

basado en la arquitectura para la integración de sistemas heterogéneos en un Java Business Integration (JBI) como propone (Vinoski 2005). Con el fin de proveer la capacidad para representar la lógica de negocios como un componente reutilizable que puede ser fácilmente integrado cuando se acopla a una aplicación, se incluyó la arquitectura de componentes de servicio basados en (Beisiegel, y otros 2005) y para manejar diferentes tipos y fuentes de datos y representarlos en una solución SOA se consideró el trabajo sobre los objetos de servicio de datos (SDO) de (Weaver 2005).

b. Descripción y arquitectura de la solución

El proceso de orquestación que describe el escenario general en términos del sistema, inicia el proceso cuando el encargado requiere insumos para el proceso de producción, realizando una solicitud del insumo. El paso siguiente en el proceso es la verificación en el almacén de insumos de la empresa; en el caso que exista el insumo, se despacha al proceso productivo, pero si no existe en el inventario se debe hacer una solicitud a un proveedor, aquí el sistema agrega una entrada en la lista de tareas del administrador de órdenes de pedido y espera hasta que el administrador de órdenes de pedido entregue la entrada (input) esperada. Cuando se recibe la entrada, el proceso continúa con la ejecución.

En este proceso de negocio simulado podemos clarificar la idea inicial del ejercicio modelada con jBPM, como se puede apreciar en la figura 1, cuando se llega al nodo tipo decisión, escoge entre dos opciones para continuar la ejecución del proceso, en la alternativa NO, tenemos tres nodos, dos nodos de tipo y uno tipo tarea, para jBPM el comportamiento interno es diferente.

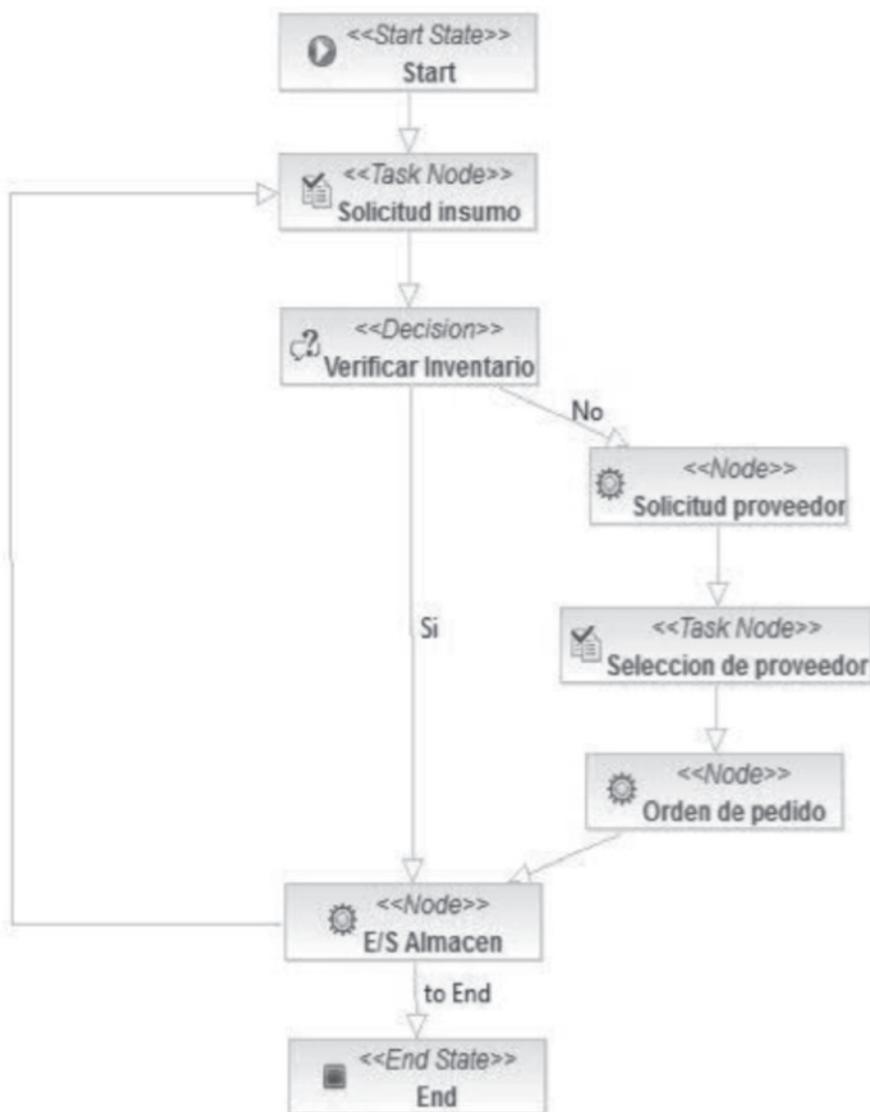


Figura 1. Proceso de negocio inicial en jpdI

Para buscar la optimización del proceso se realizó una modificación en el diseño del proceso de negocio que permitiera la integración del árbol de decisión en el de forma automática, esto se logró a través del cambio del nodo tipo tarea por el nodo de tipo, de esta forma se elimina el estado en espera y se hace el acople con la herramienta de minería de datos weka y la base de datos postgresql a través de los diferentes proyectos presentados en este trabajo.

Los datos de entrada para la evaluación del árbol de decisión son almacenados en la base de datos postgresql y entregados al algoritmo usando el formato JSON para el intercambio de datos entre la base de datos y weka. Debido a que weka utiliza una estructura de archivo denominado ARFF, se hizo necesario convertir del formato JSON a ARFF, de tal forma que los datos con los criterios para la selección de proveedores entregados en el proceso de negocio se intercambian entre diferentes tecnologías de forma transparente.

Adicionalmente, en este prototipo se incluyó un servicio web responsable de poner a disposición la gestión del proceso de negocio que será consultado por un cliente a través de una interface creada para este propósito, esta arquitectura se puede apreciar en la figura 2.

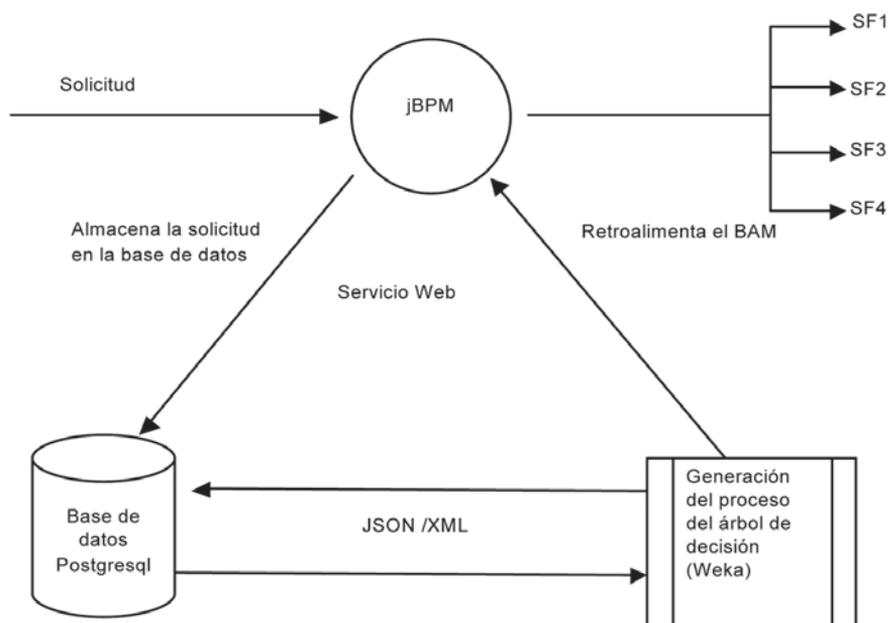


Figura 2. Arquitectura del prototipo

En el proceso de negocio diseñado en la suite de jBPM, se asoció una acción (action handler) en uno de los nodos que invoca el proceso para la evaluación de los diferentes proveedores. En este proceso se toman los datos asociados a cada posible proveedor y se almacenan en la base de datos Postgresql en la tabla movimientos; estos datos se transforman en el formato JSON con el fin de realizar el intercambio con el algoritmo del árbol de decisión; el formato de los archivos de datos es ARFF.

Después de generar el árbol de decisión se obtiene una matriz de confusión que contiene el número de fabricantes de cada clase que cumple con los criterios para ser proveedores; se reconocen los proveedores que se tienen por cada clase y se llevan nuevamente a la base de datos, se crean tantas tablas como proveedores se tengan y se almacena el fabricante y su cantidad, posteriormente, estos datos son llevados al nodo del proceso de negocio y consultados a través de un servicio web.

El siguiente diagrama muestra el flujo entre los diferentes componentes que conforman el prototipo.

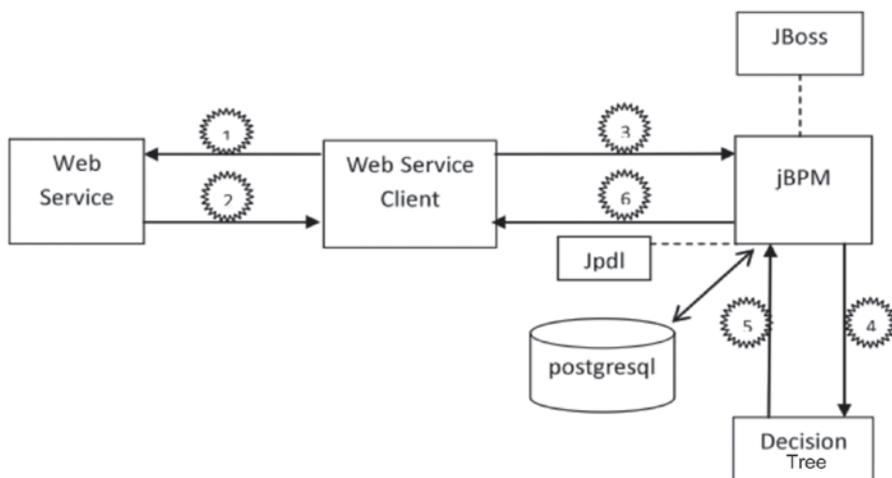


Figura 3. Diagrama de flujo del prototipo

1. En la interface de usuario se hace la solicitud de lectura al servicio web del proceso de negocio activo.
2. El servidor ejecuta la solicitud y le envía al usuario los procesos de negocio definidos.
3. El usuario inicia el proceso invocando el jBPM con el proceso de negocio que posee la etapa de selección de proveedores.
4. La información con los criterios para la selección de proveedores es procesada en el árbol de decisión usando el algoritmo J48.
5. La matriz de confusión resultante es retornada al proceso de negocio usando JSON.
6. El resultado es mostrado en la interface de usuario y almacenado en la base de datos postgresql.

c. Implementación de la solución

Para la integración de los diferentes componentes se parte del módulo de diseño de procesos de negocio de la suite de jBPM, aquí se implementan cada uno de los elementos que forman el sistema integrado. Utilizando la plataforma de diseño Eclipse, se logra la obtención de un sistema modular que integra los componentes de BPM, Minería de datos, JSON y servicio Web.

La implementación se lleva a cabo por medio de un sistema orientado a objetos, conformado por un modelo de clases organizado por paquetes y proyectos de desarrollo.

Los proyectos de desarrollo se componen de los siguientes elementos:

- jbpAnalysis
- Minería
- WebServiceProject

El proyecto jbmpAnalysis está organizado de tal forma que ejecute, mediante previo modelado, un proceso de negocio. Dicho diseño se encuentra modelado en jpdL, La figura 4 representa el caso implementado, en esta suite el archivo que contiene el proceso es processdefinition.xml.

Como estructura general, se presenta un proyecto que tiene como elemento principal la clase clsProveedor, la cual, por medio del método activaProceso() inicia la lectura del proceso de negocio para su posterior ejecución.

Como se observa en la figura 4, la clase clsProveedor usa la definición de jBPM MessageActionHandler como la relación de los mensajes emitidos entre los nodos, esta es la base fundamental para el llamado de los procedimientos para lograr la minería de datos, los cuales se llevan a cabo por medio de la clase clsMinería.

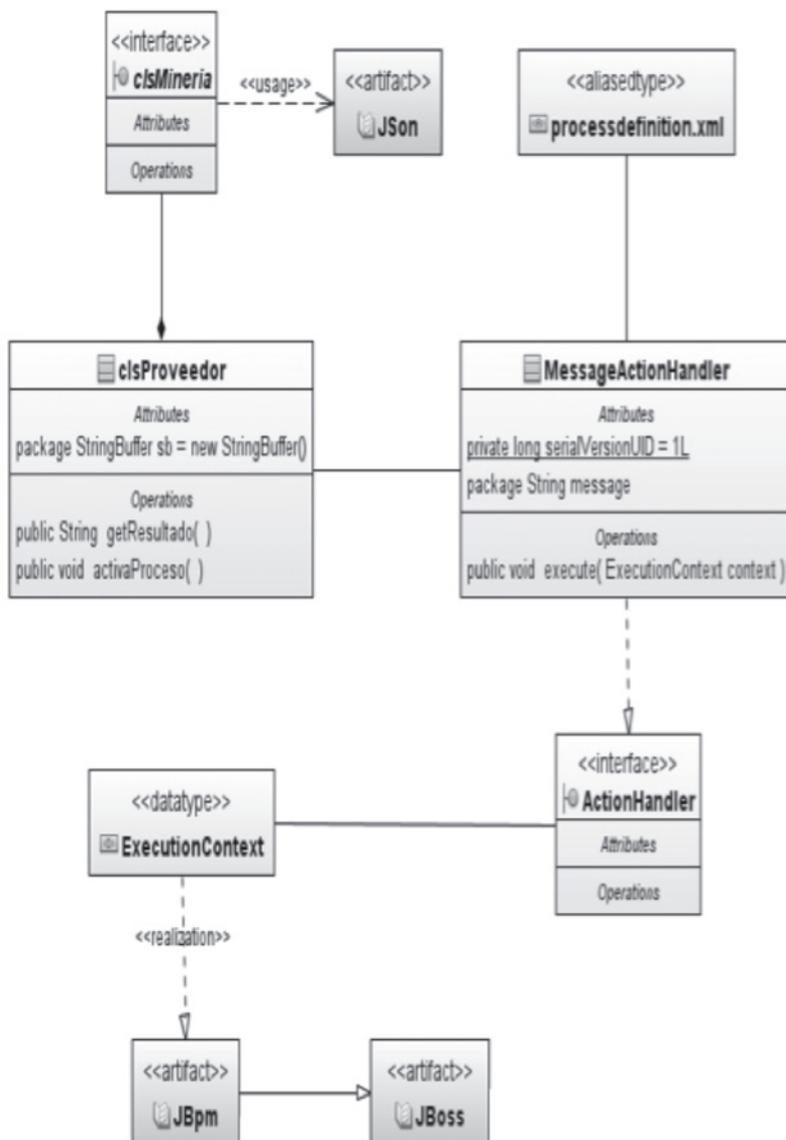


Figura 4. jbpmAnalysis Project

El proyecto Minería contiene la implementación del algoritmo J48 que trae la herramienta Weka para el análisis y selección de los proveedores, realiza el almacenamiento e intercambio de los datos con la base de datos postgresQL y se encarga de hacer la adaptación del estándar JSON al formato ARFF que requiere weka para funcionamiento.

Luego de generados los elementos de procesamientos de modelo de negocio y minería de datos, es necesario crear un proyecto que ofrezca, mediante un servicio web, los resultados solicitados por un cliente determinado, esta es la función realizada por el proyecto WebServiceProject.

3. RESULTADOS Y DISCUSIÓN

Al evaluar el árbol de decisión los registros con los posibles proveedores, devuelve una matriz de confusión con la cantidad de proveedores que cumplen con los criterios establecidos para ser proveedores. Al examinar la matriz de confusión o tabla de contingencia, el número de instancias clasificadas correctamente es la suma de los números en la diagonal de la matriz; los demás están clasificados incorrectamente. Al aplicar el método de validación se obtuvo una precisión de 66,67% con 40 instancias de las 60 clasificadas correctamente. Para el ejemplo de prueba (Use training set) se obtuvieron los siguientes valores de la matriz confusión:

```
=== Confusion Matrix ===
 a  b  c  d  <-- classified as
 8  2  2  3 | a = intel
 1 10  3  1 | b = amd
 0  1 13  1 | c = nvidia
 1  3  2  9 | d = via
```

Al evaluar la diagonal se observó que 8 posibles proveedores de intel cumplen los criterios, para amd se hallaron 10 proveedores que están clasificados correctamente, para nvidia se obtuvieron 13 proveedores clasificados correctamente y por último para via arrojó 13 proveedores.

Estos resultados se entregan a la etapa del proceso de negocio que generó la solicitud de evaluación con el árbol de decisión usando para ello un servicio web. Este proceso era realizado anteriormente con intervención humana cuyo resultado era entregado de forma manual a la etapa del proceso de negocio (selección de proveedores) que lo invocaba, ahora es un proceso completamente automático y transparente para el usuario, debido a que sólo se requiere el archivo con la información de los proveedores y la aplicación devuelve en la misma etapa del proceso de negocio el número de proveedores suministrados que cumplen los requisitos del árbol para ser considerados como una instancia correcta o incorrecta.

4. CONCLUSIONES Y TRABAJO FUTURO

Las diferentes tecnologías Open source permiten la integración de diferentes sistemas usando componentes de diferentes aplicaciones siguiendo el principio de SOA. Para este proyecto se conectó la suite de jBPM que permite en la definición de los nodos del proceso de negocio, asociarle una instancia de código Java que interactúa con sistemas externos, en este caso específico con la base de datos postgresql y el árbol de decisión extraído desde Weka, con intercambio de datos en formatos estándares (JSON). Sin embargo, para la integración de los componentes se requirió el desarrollo de clases intermedias para unir estas tecnologías.

A pesar de lograr la integración de todas las tecnologías formuladas en el trabajo, la propuesta inicial de utilizar un árbol de decisión para la selección de proveedores no es suficiente, debido a que la matriz confusión de resultados devuelve el número de proveedores que cumplen con los criterios de selección; sería recomendable pensar en realizar un híbrido con otra técnica de minería de datos para complementar y obtener el resultado esperado.

La versión de jBPM utilizada en este proyecto, aunque permite trabajar las etapas de modelado, implementación y ejecución, no posee soporte para las etapas de simulación y optimización, situación que impide desarrollar todo el ciclo de vida de BPM, además de carecer de adherencia a los estándares ni BPMN, XPDL o BPEL, soporta un modelo interno denominado JPDL, generando una brecha entre el diseño de los procesos y su implementación.

Las empresas están enfocadas a realizar diseños de sus procesos de negocio, pero es importante determinar si efectivamente el proceso diseñado responde adecuadamente a las necesidades para las cuales fue generado. Process mining permite descubrir el proceso en sí a través del registro y análisis de la información generada por eventos y determinar sus debilidades, composición, ejecución; así mismo, permite madurar y fortalecer el proceso y su gestión.

El mundo de los negocios y el mundo de las tecnologías de la información pueden relacionarse mejor con la incorporación de un procesamiento semántico, que combine la web semántica y los servicios web semánticos con BPM (Laliwala, y otros 2006) denominados sistemas SBPM.

BIBLIOGRAFÍA

- Ari, I., J. Li, A. Kozlov, y M. Dekhil. 2008. ***Data Mining Model Management to Support Real-Time Business Intelligence in Service-Oriented Architectures***. HP Software University Association Workshop.
- Ari, I., J. Li, J. Jain, y A. Kozlov. 2008. ***Management of Data Mining Model Lifecycle to Support Intelligent Business Service. ACM Computer Human Interaction for Management of IT (CHIMIT'08)***. San Diego CA.
- Beisiegel, M., y otros. 2005. ***Service Component Architecture. Building Systems using a Service Oriented Architecture***. BEA, IBM, Interface21, IONA, Oracle, SAP, Siebel, Sybase, white paper, version 0.9 edition, 3-31.
- Desmet, Stein, Bruno Volckaert, Steven Van Assche, y Dietrich Van der Weken. 2007. ***Throughput Evaluation of Different Enterprise Service Bus Approaches. Software Engineering Research & Practice***. Las Vegas, 378-384.
- Krafzig, D., K. Banke, y D. Slama. 2004. ***Enterprise SOA: Service-Oriented Architecture Best Practices (The Coad Series)***. NJ, USA: Prentice Hall PTR Upper Saddle River.
- Laliwala, Zakir, Rahul Khosla, Pritha Majumdar, Chaudhary, y Sanjay. 2006. «***Semantic and Rules Based Event-Driven Dynamic Web Services Composition for Automation of Business Processes***. IEEE Services Computing Workshops (SCW'06), 175-182.
- Lohrmann, M., y M. Reichert. 2010. ***Basic Considerations on Business Process Quality***. Ulm: University of Ulm.
- Molina, J.G, M.J Ortin, B Moros, J Nicolás, y A. Toval. 2000. ***De los Procesos del Negocio a los Casos de Uso. Jornadas de Ingeniería del Software y Bases de Datos***. Pág, 103-116.
- Rinderle-Ma, Stefanie, y Wil M.P. van der Aalst. 2007. ***Life-Cycle Support for Staff Assignment Rules in process-aware information systems***. Eindhoven: TU Eindhoven.
- Seng, Jia-Lang, y T.C. Chen. 2010. ***An Analytic Approach to***

Select Data Mining for Business Decision. Expert Systems with Applications, 8042–8057.

- Van der Aalst, W.M.P, A. H. M. Hofstede, y M. Weske. 2003. **Business Process Management: A survey.** Lecture Notes in Computer Science, 1-12.
- Vinoski, Steve. 2005. **Java Business Integration.** IEEE Internet Computing, 89-91.
- Weaver, Rick. 2005. **The Business Value of the Service Component Architecture (SCA) and Service Data Objects (SDO).** IBM Whitepaper, 1-15.