

USO DE MINERÍA DE DATOS PARA IDENTIFICAR COMPORTAMIENTOS ERRÁTICOS EN EL SISTEMA DE LIQUIDACIÓN DE HABERES DEL GOBIERNO DE LA PROVINCIA DE CATAMARCA

María A. Barrera, Carolina I. Chayle, Claudia M. Herrera, Eugenia C. Sosa Bruchman, Natalia E. Fernandez, Andrea M. Zelarayan, Silvia Baigorria

Departamento Sistemas/Facultad de Tecnología y Ciencias Aplicadas/Universidad Nacional de Catamarca
Maximio Victoria N° 55 - C.P: 4700 - San Fernando del Valle de Catamarca

Telefono: 03833- 435112 – int 168

marita.latingeo@gmail.com, cchayle@hotmail.com, cm_herrera@tecno.unca.edu.ar & sosab_ec@hotmail.com

Resumen

En la actualidad la Minería de Datos (Data Mining) permite, a partir de diversos datos, obtener información sumamente relevante. Por otra parte, la liquidación de haberes es una de las actividades más importante en cualquier empresa y/u organización, la cual está sujeta a posibles errores al momento de efectuarla. Es por ello que la motivación de este trabajo es la de encontrar todas las posibles fallas o comportamientos erráticos, en la emisión de la Liquidación de Haberes de los empleados de la Administración Pública de la Provincia de Catamarca los cuales, debido al gran volumen de datos que se manejan, son difíciles de detectar y solucionar en forma manual. A partir de lo antes expuesto, se muestra la utilidad que brindan los procesos de la Minería de Datos en la búsqueda de conocimientos ocultos (fallas, comportamientos erróneos o anómalos) en las Bases de Datos del Sistema de Liquidación de Haberes; demostrándose de esta manera que, si la Dirección de Liquidación de Haberes, del Gobierno de la Provincia de Catamarca, aplicara en un futuro dicho proceso podría minimizar fallas en el proceso de liquidación de haberes.

Palabras clave: Minería de Datos (Data Mining); Bases de Datos; Liquidación de Haberes.

Contexto

El proyecto pretende a través de la combinación de las Bases de Datos, soporte de decisiones, y Minería de Datos (Data Mining) suministrar un innovador y revolucionario enfoque de la administración de la información

utilizando nuevas tecnologías para poseer conocimientos correctos, y de esta manera, generar respuestas correctas y realizar decisiones estratégicas para la ejecución de los distintos organismos. En este caso en particular, se considera al Gobierno de la Provincia de Catamarca como una empresa, con necesidades especiales, cuantiosas y de envergadura, con grandes Bases de Datos y por consiguiente el manejo de grandes volúmenes de datos.

Así mismo es prioritario para el Departamento de Sistemas de la Facultad de Tecnología y Ciencias Aplicadas de la Universidad Nacional de Catamarca (UNCa), transferir soluciones concretas a las problemáticas que se plantean a los profesionales y a los organismos provinciales, en relación con el aseguramiento de la calidad y evaluación de los riesgos de las TI, para garantizar la seguridad de la información y prevenir las posibles contingencias en el uso de TI.

Introducción

Existen hoy en día diferentes situaciones en las cuales las empresas o instituciones ponen en juego la calidad de su trabajo, éxito y hasta el prestigio de las mismas. Esto se da como resultado del mal manejo de los datos o por la falta de conocimientos en la forma de cómo administrarlos.

En los entornos laborales lo que se busca es la mayor y mejor calidad, y seguridad de los datos, para lograr un buen desempeño de las empresas y/o instituciones. Uno de los métodos más usados actualmente para obtener información oculta es la Minería de Datos (Data Mining), una vez que obtenemos esta

información oculta, tales como fallas o diferencias que no se pueden apreciar a simple vista, lo que se logra es poder predecir o detectar estas fallas, errores o anomalías llamados también comportamientos erráticos, permitiendo la toma de decisiones proactivas. La Minería de Datos surge como una tecnología que intenta ayudar a comprender el contenido de una Base de Datos. [MOLINA et. al]

De forma general, los datos son la materia prima bruta, en el momento que el usuario les atribuye algún significado especial pasan a convertirse en información. Cuando los especialistas elaboran o encuentran un modelo, haciendo que la interpretación confronte la información, y ese modelo represente un valor agregado, entonces nos referimos al conocimiento. En la Figura N° 1 se ilustra la jerarquía que existe en una Base de Datos entre dato, información y conocimiento. Se observa igualmente el volumen que presenta en cada nivel y el valor que los responsables de las decisiones le dan a esa jerarquía. El área interna dentro de la pirámide representa los objetivos que se han propuesto. La separación de la pirámide representa la estrecha unión entre dato e información, no así entre la información y el conocimiento. La Minería de Datos trabaja en el nivel superior buscando patrones, comportamientos, agrupaciones, secuencias, tendencias o asociaciones que puedan generar algún modelo que nos permita comprender mejor el dominio para ayudar en una posible toma de decisión. [VALCARCEL ASENCIOS]

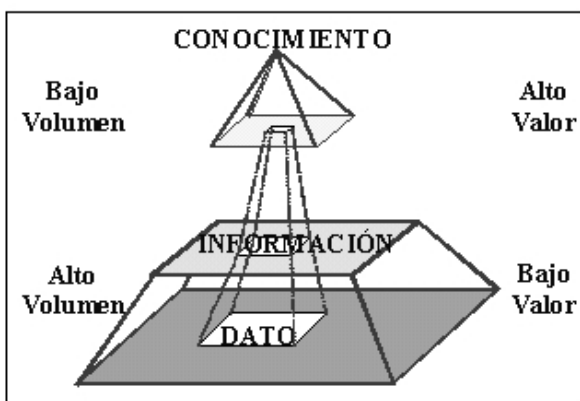


Figura N° 1 Relación entre Dato, Información y conocimiento.

Por lo tanto, la Minería de Datos o Data Mining es una herramienta que nos ayuda a analizar grandes bloques de información, mediante la explotación de los datos, sin perder el objetivo principal de cualquier empresa y/o institución. El proceso de los almacenes de datos se orienta a proveer la información correcta, a la persona indicada, en el formato adecuado y en el tiempo preciso. Su aplicación es ventajosa en cualquier área de negocios donde sea necesario mejorar el proceso de toma de decisiones, acceder a información clave, etc.

En este caso particular la investigación se centró en la Dirección de Liquidación de Haberes dependiente de la Dirección Provincial de Gestión de la Información (DPGI), del Gobierno de la Provincia de Catamarca, que dentro de sus funciones específicas es la encargada de realizar la liquidación de haberes de los empleados del sector público. El proceso de liquidación tiene varias actividades como lo es el control de todos los datos que afectan la liquidación de sueldos: cumplir con las normas legales vigentes, generar los recibos de haberes en tiempo y forma, generar el pago de haberes en tiempo y forma, generar todos los cálculos de forma totalmente automática, controlar los resultados obtenidos de la liquidación de haberes, carga y consulta de datos de la liquidación de haberes, entre otras.

Líneas de investigación y desarrollo

Dentro del área de Bases de Datos y Minería de Datos, el desarrollo de este trabajo se basó en el análisis de las Bases de Datos de la Liquidación de Haberes correspondientes al segundo semestre del año 1995, para mantener la integridad de los agentes de la Administración Pública de la Provincia de Catamarca. Como son Bases de Datos de gran volumen de aproximadamente 330.000 registros, con lo cual es imposible realizar el manejo manual de los mismos, se utilizó la Minería de Datos (Data Mining) para el procesamiento de los datos. Cabe aclarar, que si bien las bases de datos utilizadas para la aplicación de la minería de datos no son recientes, las que actualmente se utilizan para la liquidación de haberes conservan el mismo

formato por lo que, para la investigación realizada, no constituyó impedimento alguno la data de dichas bases. Para la explotación de dichos datos se utilizó la metodología CRISP-DM (CRoss- Industry Standard Process for Data Mining) [GALLARDO ARANCIBIA] para dar cumplimiento con las fases de la Minería de Datos. A través del uso de la técnica Clusterización, también llamada Agrupamiento, y de la herramienta de libre distribución WEKA (Waikato Environment for Knowledge Analysis) [WEKA] se obtuvo la información necesaria de los datos para su posterior análisis.

La metodología CRISP-DM, se utilizó en sus cuatro niveles de abstracción: fases, tarea genérica, tarea especializada e instancia de proceso; los cuales se encuentran organizados de forma jerárquica que van desde el nivel más general hasta los más específicos. [GALLARDO ARANCIBIA]

La búsqueda de información oculta estuvo orientada a la detección de diferentes errores que surgen en proceso de liquidación de haberes; posteriormente, se confeccionó un informe final con los resultados obtenidos, brindando así a la Dirección de Liquidación de Haberes información útil para la posterior toma de decisiones respecto de la aplicación e implementación futura del proceso de Minería de Datos a fin de detectar y dar solución a las situaciones erróneas que presenta el Sistema de Liquidación de Haberes.

Resultados y Objetivos

El objetivo principal del presente proyecto fue aplicar Minería de Datos (Data Mining) para identificar los comportamientos erráticos en el proceso de Liquidación de Haberes de la Administración Pública de la Provincia de Catamarca.

Es importante destacar que todo el proceso de desarrollo estuvo guiado por las pautas definidas en la metodología CRIPS-DM, lo cual facilitó el alcance de los objetivos definidos, y fue implementado en la herramienta de libre distribución WEKA. Esto demuestra que es factible aplicar tecnologías de bajo costo de implementación.

Luego de la evaluación de las Bases de Datos de la Liquidación de Haberes,

correspondientes al segundo semestre del año 1995, se realizó un resumen estadístico de los errores encontrados los cuales se detallan en la siguiente tabla:

Tipo de Error	Cantidad de Veces Encontrado	Porcentaje
Liquidación docentes hora cátedra	53	39%
Liquidación por decretos no especificados	46	34%
Liquidación en cero	30	22%
Liquidación Incorrecta sin Justificación	7	5%

Resumen Estadístico de Errores Encontrados

Durante el desarrollo del proyecto se plantearon algunos problemas, los que se enumeran a continuación:

- Las bases de datos se encontraban en el lenguaje COBOL. El problema de que estaban codificadas en este lenguaje fue que la herramienta WEKA no reconocía este tipo de formato.
- Bases de datos de gran tamaño. El problema que se presentó fue que eran demasiado grande las bases de datos para poder trabajarlas con la herramienta WEKA.
- Falta de información. La información no estaba almacenada o se guardaba en forma incorrecta, además se detectaron archivos borrados.

Una vez analizado cada caso que se presentó se puede decir que las acciones a tomar para minimizar los errores son:

- Agregar información con más detalle dentro de las bases de datos.
- Ampliar las validaciones para evitar que se generen liquidaciones con importes en cero.
- En los casos en que se realice la liquidación por decreto, especificar en cada caso número de decreto con el que se paga al agente.
- Registrar debidamente las bajas y altas de los agentes a los que se les realice la liquidación de haberes.

El proyecto ha demostrado la utilidad que brindan los procesos de Minería de Datos,

junto a sus técnicas y herramientas, en la búsqueda de conocimientos ocultos en grandes bases de datos, en este caso en particular, la obtención de comportamientos erráticos, lo cual ayudará a mejorar el proceso de la Liquidación de Haberes realizado a los empleados de la Administración Pública de la provincia de Catamarca.

Formación de Recursos Humanos

Las integrantes Lic. Andrea M. Zelarayán y Silvia Baigorria desarrollaron como tesina de grado, en el marco de la carrera de Licenciatura en Sistemas de Información, el tema del presente proyecto específicamente relacionado al área de Bases de Datos y Minería de Datos aplicado en el gobierno de la provincia de Catamarca, bajo la dirección de la Lic. Claudia M. Herrera.

Además, la Lic. María A. Barrera, la Lic. Claudia M. Herrera y la Lic. Eugenia C. Sosa Bruchman son docentes de la cátedra de “Auditoría de Sistemas” de la carrera Ingeniería en Informática de la Facultad de Tecnología y Ciencias Aplicadas; y la Lic. Carolina Irene Chayle es docente de la cátedra de “Organización y Sistemas” de la carrera Ingeniería en Informática que se dicta en dicha Unidad Académica, estando ambas cátedras estrechamente ligadas al tema en cuestión.

Específicamente el proyecto de investigación prevé el programa de capacitación y formación de recursos humanos, que contempla las siguientes actividades:

- Incorporación de alumnos de los últimos años de la carrera de Ingeniería en Informática de la Facultad de Tecnología y Ciencias Aplicadas de la UNCa., en calidad de auxiliares de investigación.
- Dirección de tesinas de grado de la carrera de Ingeniería en Informática de la Facultad de Tecnología y Ciencias Aplicadas de la UNCa.
- Participación de los integrantes del proyecto en cursos de actualización y posgrado en el área de estudio.
- Participación en talleres o workshops de herramientas informáticas relacionadas con el control y auditoría de TI.

- Celebración de convenios con entes estatales, como la Dirección Provincial de Gestión de la Información de la Provincia de Catamarca, la Municipalidad de la Capital y otras universidades, como así también con el sector privado para capacitación de los recursos humanos en las herramientas normalizadas.
- Dictado de cursos de capacitación diseñados según demanda de la propia Administración, con el objeto de transmitir las mejores prácticas a los agentes públicos asistentes.

Para garantizar la capacitación y actualización del equipo de investigación, así como la difusión de los avances y resultados logrados, se propuso la participación en eventos nacionales e internacionales de la especialidad, como congresos, simposios, seminarios y cursos.

Bibliografía

GALLARDO ARANCIBIA José Alberto, *Metodología para el Desarrollo de Proyectos en Minería de Datos CRIPS-DM*, Facultad de Informática, Universidad Politécnica de Madrid.

MOLINA L.C. and Sangüesa, *Data Mining*, Universitat Oberta de Catalunya.

VALCARCEL ASENCIOS Violeta, *Data Mining y el descubrimiento del conocimiento*, Industrial Data – Revista de Investigación, Universidad Nacional Mayor de San Marcos, Perú.

WEKA, *Waikato Environment for Knowledge Analysis*, The University of Waikato.