Análisis de los Datos Históricos de la Programación de Cursos en los CECATI del Estado de Colima

Historical Data Analysis for Scheduling of Cecati's Courses in Colima State

Manuel Espinosa Ortega
Instituto Tecnológico de Colima,
g7946192@itcolima.edu.mx

Nicandro Farías Mendoza
Instituto Tecnológico de Colima,
nmendoza@ucol.mx

Jesús Alberto Verduzco Ramírez
Instituto Tecnológico de Colima,
averduzco@itcolima.edu.mx

Resumen

Hoy en día las herramientas para la extracción de información están mejorando el proceso para que las empresas y dependencias puedan obtener información a partir de grandes volúmenes de datos. Los sistemas de extracción de información se aplican tradicionalmente como una secuencia de módulos de propósito especial, la extracción se convierte, como una clase particular de piezas relevantes de información, que son utilizados por las dependencias o empresas con el fin de tomar decisiones que mejoran la funcionalidad de sus procesos. En este documento se hace una descripción general del Sistema Web para la Programación de Cursos en los CECATI (SWPCC). En particular, nuestra investigación se enfoca a desarrollar un módulo para la extracción de información, a partir del análisis de datos históricos, de la programación de cursos en los CECATI del Estado de Colima, durante el ciclo escolar 2013-2014 mediante herramientas de Data Warehouse y Minería de Datos. El análisis de los datos históricos arroja información sobre los cursos más

programados, los escasamente programados, así como otras áreas de oportunidad y otros aspectos como los horarios y duración de los cursos que pueden influir en la demanda de los cursos que se imparten en los CECATI, lo que nos permite tomar las decisiones para lograr una planeación efectiva de los cursos. La metodología empleada para nuestra investigación, consiste en las siguientes tres fases:

En la primera fase se presenta la contextualización de este trabajo de investigación, describiendo los mecanismos o técnicas de extracción de información, la exposición de las bases de datos históricas, el Sistema Web para la Programación de Cursos en los CECATI (SWPCC) y el módulo de extracción de información.

En una segunda fase presenta en forma detallada cada una de las etapas que se realizaron para obtener un set de datos limpios que se pudieran analizar mediante una herramienta llamada Weka (Waikato Environment for Knowledge Analysis) (Sudhir, Kodge, 2013) a partir de un conjunto de datos de origen de los CECATI en el Estado de Colima del ciclo escolar 2013-2014, la conversión del formato de este archivo original, el procesado de los datos, el análisis de los datos, el trabajo de filtrado, y la discretización de los datos.

En una tercera fase se detallan cada uno de los resultados que se obtienen en el análisis de los datos con el uso de los diferentes algoritmos que posee Weka, procurando una presentación en forma sencilla y clara de estos resultados, de manera que puedan brindar a los interesados nueva información para la toma de decisiones. La metodología de trabajo que se detalla en este documento puede servir de base para futuras investigaciones con otros ciclos escolares, con propósitos de obtener nuevos conocimientos.

Key words: Minería de datos, Data Warehouse, datos históricos, proceso KDD.

Abstract

Today the tools for information extraction are improving the process for companies and dependencies can obtain information from large volumes of data. The information extraction systems are traditionally implemented as a sequence of special-purpose modules, extraction is converted, as a particular class of relevant pieces of information, which they are used by dependencies or companies in order to make decisions that enhance the functionality of its processes. This document provides an overview of the Web System for

Programming of Courses in the CECATI (SWPCC). In particular, our research focuses on developing a module for extracting information, from the analysis of historical data, of programming of courses in the CECATI of state of Colima, during the 2013-2014 school year, using the tools of Data Warehouse and Data Mining. The analysis of historical data yields information about the more scheduled courses, the less programmed, as well as other areas of opportunity and other aspects such as the time and duration of the courses that can influence the demand for courses taught in CECATI, allowing us to make decisions for effective planning of courses. The methodology used for our research, consists of the following three phases:

In the first phase the contextualization of this research is presented, describing the mechanisms and techniques of information extraction, the exposure of historical databases, Web System for Programming of Courses in CECATI (SWPCC) and module of information extraction.

In a second phase presents in detail each of the steps that were performed to obtain a set of clean data that could be analyzed by a tool called Weka (Waikato Environment for Knowledge Analysis) (Sudhir, Kodge, 2013) from a dataset origin of CECATI in the State of Colima during the 2013-2014 school year, the conversion of the original file format, the data processing, the data analysis, the data filtering and discretization the data.

In a third phase detailed each of the results obtained in the analysis of data using different algorithms that has Weka, endeavoring a presentation in simple and clear way of these results, so that can provide to interested parties new information for decision-making. The methodology that is detailed in this document can serve as a basis for future research with other school years, with the purpose of obtaining new knowledge.

Key words: Data Mining, Data Warehouse, historical data, KDD process.

Fecha Recepción: Mayo 2015 **Fecha Aceptación:** Enero 2016

Introducción

Muchas empresas no le dan la importancia que se debería al uso de la tecnología por lo que es de suma importancia para el desarrollo de las mismas ya que se requiere hoy en día empresas que compitan en el mercado electrónico y que tengan sistemas de información adecuado a sus necesidades (Talo, 2015). Hoy en día las empresas deben de contar con (SI) sistemas de información que les permita obtener información confiable y que les ayude a la toma de decisiones (Rojas, 2010), de hecho existe la afirmación de que quien controle la información dominará al mundo (Johnson, 2011) y es evidente que en la actualidad existen enormes cantidades de información, por tal motivo para controlar esos enormes volúmenes de información es necesario del uso de herramientas que permitan poder extraer, ordenar, clasificar, limpiar, estructurar, transformar, etc. de manera que dicha información se convierta en útil para la toma de decisiones. Por esta necesidad surgieron mecanismos como Data Warehouse y de Minería de Datos que permiten realizar dichas tareas en beneficio de las Instituciones.

Data Warehouse

Tras las dificultades de los sistemas tradicionales en satisfacer las necesidades informacionales, surge el concepto de Data Warehouse, como solución a las necesidades informacionales globales de la empresa (Boza, 2004 et al). El término de Data Warehouse fue acuñado por primera vez por Inmon (Pablos,Albarrán, Castilla, 1998), se traduce literalmente como Almacén de Datos. No obstante si el Data Warehouse fuese exclusivamente un almacén de datos, los problemas seguirían siendo los mismos que en los Centros de Información[1]. La ventaja principal de este tipo de sistemas se basa en su concepto fundamental, la estructura de la información. Este concepto significa el almacenamiento de información homogénea y fiable, en una estructura basada en la consulta y el tratamiento jerarquizado de la misma, y en un entorno diferenciado de los

[1]Son centros especializados, creados con el propósito de recopilar datos, producir información y ponerla al alcance de todas aquellas instituciones, universidades, gremios y asociaciones empresariales, así como para la cooperación internacional. Rosanna Silva "Centros de Información y Documentación (México) www.monografias.com

sistemas operacionales. Según definió Inmon (2000), el Data Warehouse se caracteriza por ser:

- Integrado: los datos almacenados en el Data Warehouse deben integrarse en una estructura consistente, por lo que las inconsistencias existentes entre los diversos sistemas operacionales deben ser eliminadas. La información suele estructurarse también en distintos niveles de detalle para adecuarse a las distintas necesidades de los usuarios.
- Temático: sólo los datos necesarios para el proceso de generación del conocimiento del negocio se integran desde el entorno operacional. Los datos se organizan por temas para facilitar su acceso y entendimiento por parte de los usuarios finales. Por ejemplo, todos los datos sobre clientes pueden ser consolidados en una única tabla del Data Warehouse. De esta forma, las peticiones de información sobre clientes serán más fáciles de responder dado que toda la información reside en el mismo lugar.
- Histórico: el tiempo es parte implícita de la información contenida en un Data Warehouse. En los sistemas operacionales, los datos siempre reflejan el estado de la actividad del negocio en el momento presente. Por el contrario, la información almacenada en el Data Warehouse sirve, entre otras cosas, para realizar análisis de tendencias. Por lo tanto, el Data Warehouse se carga con los distintos valores que toma una variable en el tiempo para permitir comparaciones.
- No volátil: el almacén de información de un Data Warehouse existe para ser leído, y no modificado. La información es por tanto permanente, significando la actualización del Data Warehouse la incorporación de los últimos valores que tomaron las distintas variables contenidas en él sin ningún tipo de acción sobre lo que ya existía.

Minería de Datos

Silberschatz, Abraham, Korth, Henry F. y Sudarshan, S. (2002) definen la Minería de Datos (MD) como un proceso automático o semiautomático que busca descubrir patrones ocultos en un conjunto de datos y que además, sean potencialmente útiles para los usuarios de la Base de Datos (BD). En la MD se contemplan diversas estrategias para identificar diferentes tipos de patrones, como son árboles de clasificación, redes neuronales, redes bayesianas, técnicas de asociación, entre otros (Olmos, González, 2007). El objetivo en todo proceso de MD es obtener patrones de interés para el usuario final. Para lograrlo, es necesario preparar correctamente a los datos para procesarlos, elegir un método adecuado para extraer los patrones deseados y finalmente, determinar cómo evaluar los patrones encontrados.

La Extracción de conocimiento está principalmente relacionado con el proceso de descubrimiento conocido como Knowledge Discovery in Databases (KDD). Es un proceso que extrae información de calidad que puede usarse para dibujar conclusiones basadas en relaciones o modelos dentro de los datos (WebMining, 2011). La siguiente figura ilustra las etapas del proceso KDD (Figura 1):

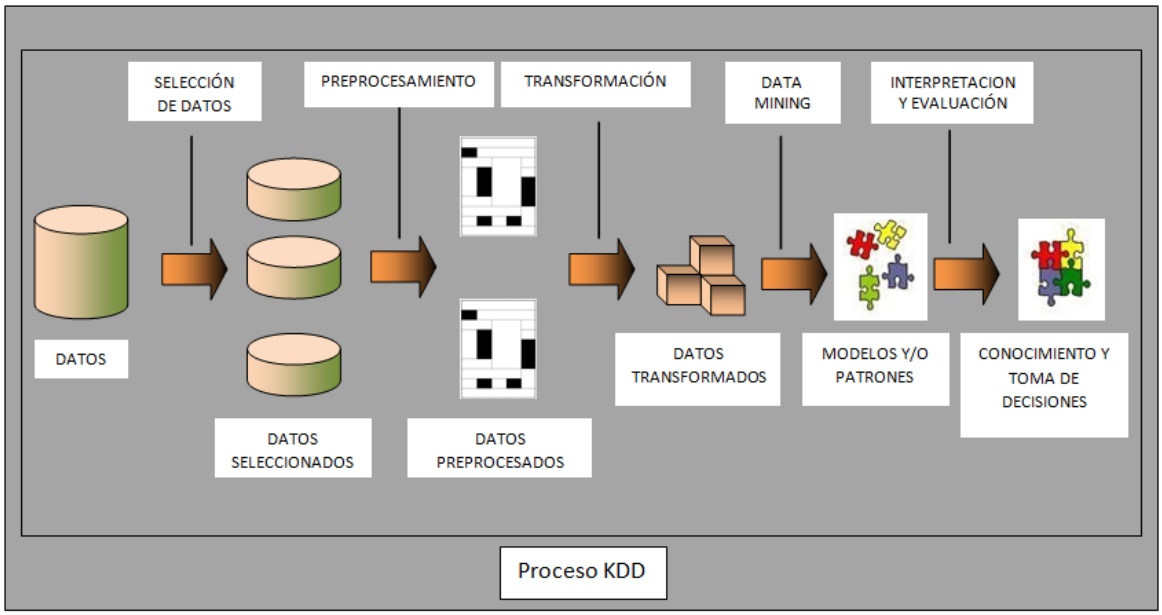

Figura 1. Etapas del proceso KDD

Como muestra la figura anterior, las etapas del proceso KDD se dividen en 5 fases y son:

Selección de datos

En esta etapa se determinan las fuentes de datos y el tipo de información a utilizar. Es la etapa donde los datos relevantes para el análisis son extraídos desde la o las fuentes de datos.

Preprocesamiento

Esta etapa consiste en la preparación y limpieza de los datos extraídos desde las distintas fuentes de datos en una forma manejable, necesaria para las fases posteriores. En esta etapa se utilizan diversas estrategias para manejar datos faltantes o en blanco, datos inconsistentes o que están fuera de rango, obteniéndose al final una estructura de datos adecuada para su posterior transformación.

Transformación

Consiste en el tratamiento preliminar de los datos, transformación y generación de nuevas variables a partir de las ya existentes con una estructura de datos apropiada. Aquí se realizan operaciones de agregación o normalización, consolidando los datos de una forma necesaria para la fase siguiente.

Data Mining

Es la fase de modelamiento propiamente tal, en donde métodos inteligentes son aplicados con el objetivo de extraer patrones previamente desconocidos, válidos, nuevos, potencialmente útiles y comprensibles y que están contenidos u "ocultos" en los datos.

Interpretación y Evaluación

Se identifican los patrones obtenidos y que son realmente interesantes, basándose en algunas medidas y se realiza una evaluación de los resultados obtenidos.

Bases de Datos Históricas

Los Centros de Capacitación para el Trabajo Industrial (CECATI) cuentan con un sistema llamado "Sistema Web para la Programación de Cursos en los CECATI" (SWPCC).Con este sistema (figura 2) se realiza en forma automática la calendarización de los cursos que se imparten en los CECATI en cada ciclo escolar.

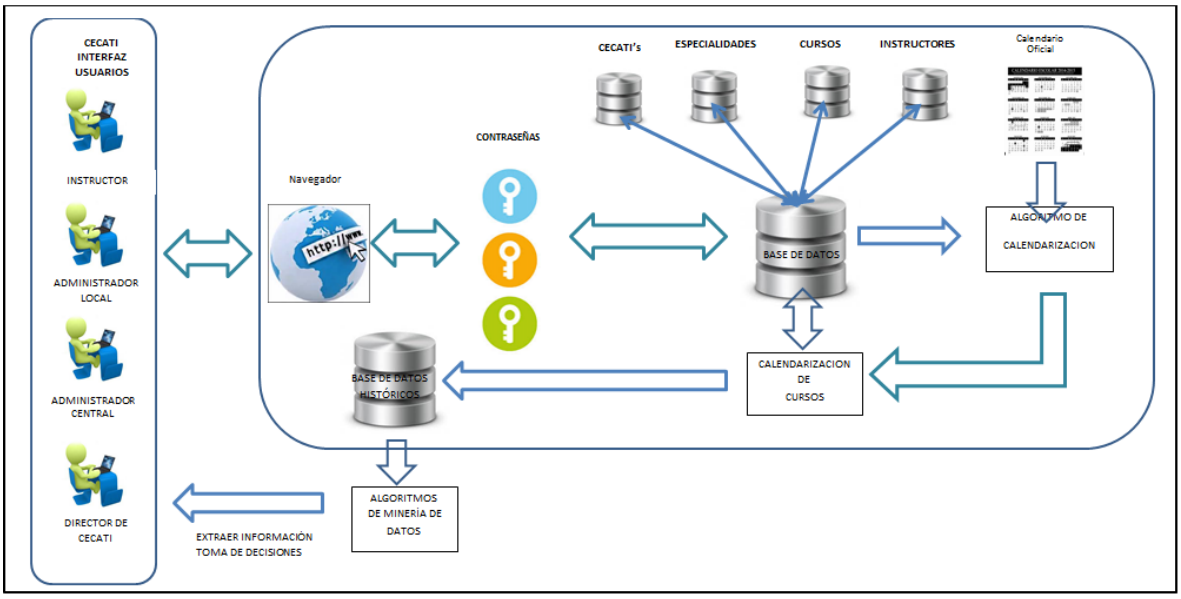

Figura 2. Sistema Web para la Programación de Cursos en los CECATI

El sistema cuenta además con un módulo de extracción de información constituido por los siguientes componentes: Primero, la base de datos histórica de las planeaciones de cursos. Esta base de datos se genera con la alimentación de la programación de los cursos, que se ofertan a nivel nacional año con año en cada CECATI. Segundo, un proceso de aplicación de algoritmos para la extracción de la información y tercero, una interface para interpretación y visualización de la información extraída (figura 3).

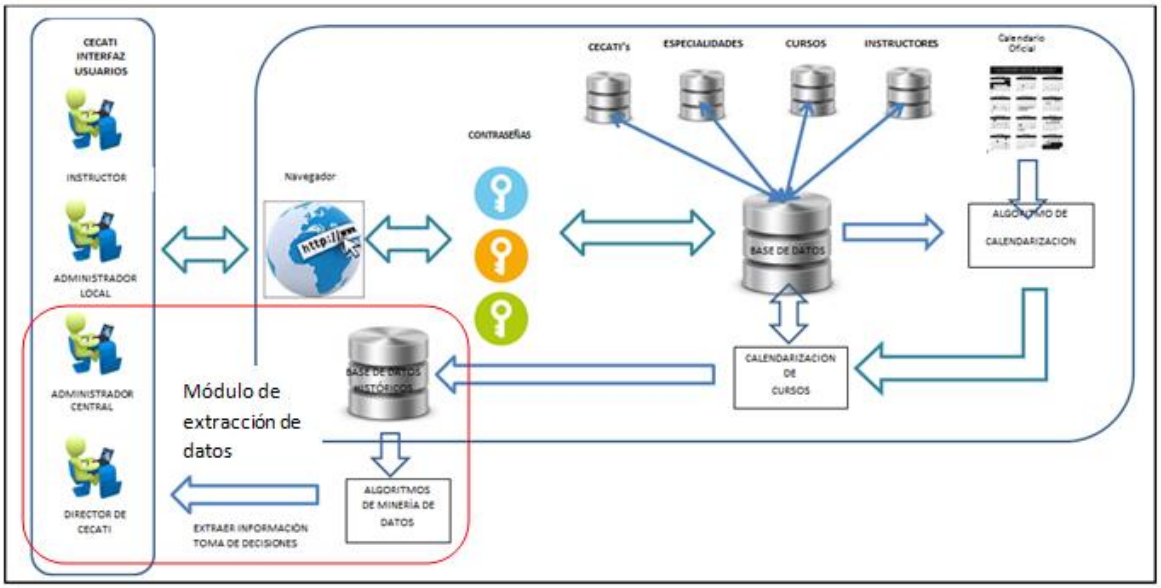

Figura 3. Módulo de extracción de información

Esta base histórica permite analizar la información mediante herramientas de Data Warehouse y Minería de Datos con propósitos de obtener patrones, tendencias o proyección de estadísticas que ayuden a la toma de decisiones (Molina, García, 2006). En este trabajo de investigación se eligió Weka por ser una aplicación de libre licencia, muy confiable en los resultados y porque posee la mayoría de los algoritmos utilizados (árboles de clasificación, redes neuronales, redes bayesianas, técnicas de asociación, etc.) para el análisis de datos. Para la realización del análisis de los datos aquí mencionados, se toman solo en cuenta los generados en los CECATI del Estado de Colima, por las facilidades de acceso a dicha información.

Obtención de los Datos

La base de datos histórica del sistema SWPCC constituye la fuente para la obtención de los datos de origen para analizar mediante Weka. El formato ".xls" corresponde a archivos que se pueden abrir con Excel, los datos de origen tienen formato ".sql", se tienen que importar con el programa Excel para cambiar el formato original ".sql" a formato ".xls". La razón de esta conversión es que Excel facilita la limpieza de los datos, dejando únicamente los datos que sean útiles para el análisis a realizar usandoWeka.

Limpieza de los Datos

Por medio de la herramienta de reemplazo de Excel se unifica el nombre de los campos para todos los atributos que tuvieran mismos datos escritos en distinta forma. Además de unificar el contenido de los datos, se eliminan los atributos que no son necesarios para el análisis del set de datos, por ejemplo: los atributos "DIRECCION", "COLONIA", "CODIGO_POSTAL", "TELEFONO", "NOMBRE_JEFE_AREA", sus contenidos son innecesarios para la información que se busca obtener; el atributo "ESPECIALIDADES" se elimina porque se puede deducir del nombre de los cursos programados; igualmente "DIAS INHABILES" también se elimina porque no es relevante su contenido.

Discretización de los Datos

El set de datos a analizar contiene algunos campos que son de tipo numérico, por ejemplo: "DURACION_EN_HORAS"; para el proceso de análisis con Weka se requiere que sean de tipo simbólico por lo tanto es necesario realizar la discretización de este tipo de datos, antes de esta acción conviene primeramente convertir el set de datos limpios al tipo

de formato CSV delimitado con comas, luego copiar los atributos y los datos en un archivo de tipo .arff (Diego García) siguiendo la siguiente estructura (Figura 4):

```
% comentarios

@relation NOMBRE_RELACION

@attribute r1 real
@attribute r2 real ...
...
@attribute i1 integer
@attribute i2 integer
…
@attribute s1 {v1_s1, v2_s1,…vn_s1}
@attribute s2 {v1_s1, v2_s1,…vn_s1}
…
@data
```

Figura 4. Estructura Weka para análisis de datos

En la parte de arriba se colocan los comentarios que sean pertinentes precedidas por el signo %. Enseguida se escribe el nombre de la relación precedida por "@relación". Posteriormente se relacionan en forma de columna cada uno de los atributos precedidos por "@attribute" seguido por el tipo de dato: real, integer, o string, según sea el caso. Esta estructura es la reconocida por WEKA y se deberá ser muy escrupuloso en la sintaxis de cada concepto ya que cualquier cambio en la misma no permitirá abrir el archivo a analizar. Debajo de @data van todos los datos delimitados por comas, tal y como se ven en el siguiente ejemplo (Figura 5):

```
%CECATI_No,NOMBRE_CURSO,TIPO_CURSO,DURACION,HORARIO_CURSO


@relationprogramacion_cursos_cecati_colima

@attribute CECATI_No {CECATI_34,CECATI_126,CECATI_145,CECATI_183}
@attribute NOMBRE_CURSO String
@attribute TIPO_CURSO {REGULAR,CAE,EXTENSION,ACCION_MOVIL}
@attribute DURACION Real
@attribute HORARIO_CURSO String
@data


CECATI_34,"INSTALACION DEL SISTEMA ELECTRICO RESIDENCIAL",REGULAR,240,"07:30-10:46"
CECATI_34,"MANTENIMIENTO DE APARATOS DOMESTICOS",REGULAR,200,"10:46-13:18"
CECATI_34,"MANTENIMIENTO DE APARATOS DOMESTICOS",REGULAR,200,"16:00-19:00"
CECATI_34,"INSTALACION DEL SISTEMA ELECTRICO INDUSTRIAL",REGULAR,240,"18:24-21:00"
CECATI_34,"INSTALACION DEL SISTEMA ELECTRICO RESIDENCIAL",REGULAR,240,"19:00-22:00"
CECATI_34,"BOBINADO DE MOTORES ELECTRICOS",CAE,57,"10:00-13:00"
CECATI_34,"INSTALACION DEL SISTEMA ELECTRICO RESIDENCIAL",REGULAR,240,"16:00-19:00"
CECATI_34,"INSTALACION DEL SISTEMA ELECTRICO INDUSTRIAL",REGULAR,240,"19:00-22:00"
CECATI_34,"MANTENIMIENTO DE SISTEMA DE A/A Y REFRIGERACION",ACCION_MOVIL,200,"10:00-16:53"
CECATI_34,"REPARACION DE REFRIGERADORES DOMESTICOS SIN ESCARCHA",EXTENSION,112,"07:00-10:00"
CECATI_126,"MANTENIMIENTO DE EQUIPOS RECEPTORES DE TELEVISION",REGULAR,360,"08:00-14:00"
CECATI_126,"REPARACIONES BASICAS DE UN AUTOESTEREO",EXTENSION,168,"08:00-14:00"
CECATI_126,"REPARACION DE MOTORES A GASOLINA",REGULAR,450,"14:00-20:00"
CECATI_126,"REPARACION DEL SISTEMA DE FRENOS BASICOS",REGULAR,280,"14:00-17:00"
CECATI_126,"REPARACION DEL SISTEMA DE FRENOS BASICOS",REGULAR,280,"17:00-20:00"
CECATI_126,"REPARACION DEL SISTEMA DE TRANSMISION MANUAL",REGULAR,234,"14:00-20:00"
CECATI_126,"CONFECCION DE PRENDAS PARA DAMA Y NIÑA",REGULAR,350,"08:00-11:00"
CECATI_126,"CONFECCION DE PRENDAS PARA DAMA Y NIÑA",REGULAR,350,"11:00-14:00"
CECATI_126,"CONFECCION DE PRENDAS PARA CABALLERO Y NIÑO",REGULAR,273,"08:00-11:00"
CECATI_126,"CONFECCION DE PRENDAS PARA CABALLERO Y NIÑO",REGULAR,273,"11:00-14:00"
CECATI_126,"ALTA COSTURA",REGULAR,350,"16:30-19:30"
CECATI_126,"ELABORACION DE BLANCOS",EXTENSION,150,"16:30-19:30"
CECATI_126,"CONFECCION DE PRENDAS PARA CABALLERO Y NIÑO",REGULAR,300,"08:00-11:00"
CECATI_126,"CONFECCION DE PRENDAS PARA CABALLERO Y NIÑO",REGULAR,300,"11:00-14:00"
CECATI_126,"ALTA COSTURA",REGULAR,350,"08:00-12:00"
CECATI_126,"DECORACION DE PRENDAS DE VESTIR",EXTENSION,176,"12:00-14:00"
CECATI_126,"DECORACION DE PRENDAS DE VESTIR",EXTENSION,120,"08:00-14:00"
CECATI_126,"INGLES COMUNICATIVO BASICO INICIAL",ACCION_MOVIL,180,"08:30-14:30"
CECATI_126,"INGLES COMUNICATIVO BASICO INICIAL",ACCION_MOVIL,180,"08:30-14:30"
CECATI_126,"VERBOS REGULARES E IRREGULARES",EXTENSION,78,"08:30-14:30"
CECATI_126,"VERBOS REGULARES E IRREGULARES",EXTENSION,66,"08:30-14:30"
CECATI_145,"MANTENIMIENTO DE AIRE ACONDICIONADO MINISPLIT",EXTENSION,111,"15:00-18:00"
CECATI_145,"MANTENIMIENTO DE AIRE ACONDICIONADO MINISPLIT",EXTENSION,111,"18:00-21:00"
```

```
CECATI_145,"INSTALACION DEL SISTEMA ELECTRICO RESIDENCIAL",REGULAR,240,"07:00-10:00"
CECATI_145,"INSTALACION DEL SISTEMA ELECTRICO RESIDENCIAL",REGULAR,240,"10:00-13:00"
CECATI_145,"INSTALACION DEL SISTEMA ELECTRICO RESIDENCIAL",REGULAR,240,"07:00-10:00"
CECATI_145,"INSTALACION DEL SISTEMA ELECTRICO RESIDENCIAL",REGULAR,240,"10:00-13:00"
CECATI_145,"INSTALACION Y REPARACION DE SISTEMAS DE COMUNICACIÓN",REGULAR,192,"07:00-11:00"
CECATI_183,"PREPARACION DE ALIMENTOS",REGULAR,360,"08:00-11:00"
CECATI_183,"PREPARACION DE BEBIDAS",REGULAR,200,"11:00-14:00"
CECATI_183,"SERVICIO A COMENSALES",REGULAR,216,"08:00-11:00"
CECATI_183,"ELABORACION DE PASTELES Y PRODUCTOS DE REPOSTERIA",REGULAR,350,"11:00-14:00"
CECATI_183,"BOCADILLOS Y PANADERIA CASERA",EXTENSION,60,"11:00-14:00"
CECATI_183,"PREPARACION DE ALIMENTOS",REGULAR,360,"16:00-20:00"
CECATI_183,"ELABORACION DE PASTELES Y PRODUCTOS DE REPOSTERIA",REGULAR,350,"16:00-20:00"
CECATI_183,"GELATINAS ARTISTICAS",EXTENSION,92,"16:00-20:00"
CECATI_183,"CORTE Y PEINADO DEL CABELLO",REGULAR,200,"08:00-11:00"
CECATI_183,"CORTE Y PEINADO DEL CABELLO",REGULAR,200,"11:00-14:00"
CECATI_183,"COLOR Y TRANSFORMACION EN EL CABELLO",REGULAR,220,"08:00-11:00"
CECATI_183,"COLOR Y TRANSFORMACION EN EL CABELLO",REGULAR,220,"11:00-14:00"
…
```

Figura 5. Estructura del análisis de datos históricos de los CECATI

Cuidando esta estructura y la sintaxis correcta, el programa de Weka (que es muy escrupuloso) recibe la información sin ninguna restricción permitiendo utilizar todas las funciones disponibles para analizar el set de datos.

Análisis de los datos

El set de datos de la Programación de Cursos en los CECATI en el Estado de Colima que en este documento se va a analizar mediante la herramienta Weka contiene los siguientes atributos: CECATI_No, NOMBRE_CURSO, TIPO_CURSO, DURACION, HORARIO_CURSO, con los que se pretende obtener información relevante de los cursos que se programaron en los CECATI en el Estado de Colima durante el ciclo escolar 2013-2014. Para el análisis del set de datos se siguen algunas recomendaciones: Los atributos cuyo contenido sea una lista de opciones bastante grande y además se requiera contengan más de una sola palabra, conviene que sean de tipo String como el caso de "NOMBRE_CURSO" se deberá remover los demás atributos y dejar solo el atributo que se va analizar, luego seleccionar el filtro **StringToNominal** para que Weka pueda desplegar la lista de instancias que contiene el atributo "NOMBRE_CURSO". Tal es el caso del atributo "HORARIO_CURSO" que es del tipo String, por lo tanto para su análisis

se procede de igual forma. A continuación se muestran las figuras de cada uno de estos atributos y de los resultados que se obtienen al aplicar el filtro **StringToNominal:**

La figura resultante (Figura 6) nos muestra un total de 111 diferentes cursos que se programaron durante el ciclo escolar 2013-2014 en los CECATI en el Estado de Colima, de los cuales, la especialidad de Servicios de belleza con los cursos de CORTE Y PEINADO DEL CABELLO, MAQUILLAJE DEL ROSTRO y COLOR Y TRANSFORMACIÓN EN EL CABELLO, son los que se programan con mayor frecuencia en relación a los demás cursos de las otras especialidades. Después de esta especialidad, la especialidad de Refrigeración y A/A con el curso MANTENIMIENTO DE SISTEMAS DE A/A Y REFRIGERACION, muestra en segundo lugar la frecuencia de programar este curso, la especialidad de Electricidad es en tercer lugar con el curso denominado INSTALACION DEL SISTEMA ELECTRICO RESIDENCIAL. Entre los cursos regulares que solamente se programaron una sola vez en todo el Estado de Colima en estos Centros de Capacitación se encuentran: MANTENIMIENTO A MOTORES ELECTRICOS, REPARACION DEL SISTEMA DE EMBRAGUE, REPARACION DEL SISTEMA DE CONTROL DE EMISION DE GASES CONTAMINANTES, REPARACION DE FRENOS ABS, MANTENIMIENTO A REDES DE AREA LOCAL (LAN), PREPARACION DE BEBIDAS Y SERVICIO A COMENSALES.

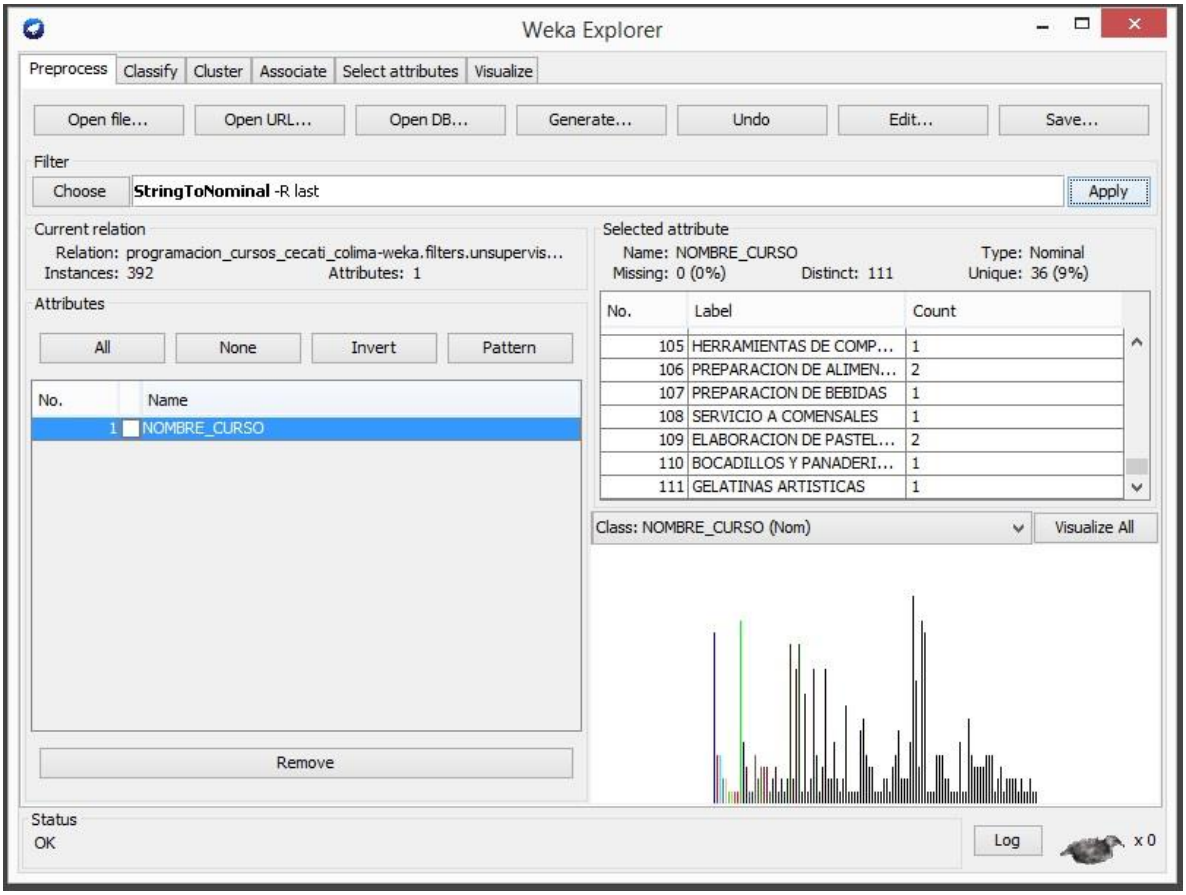

Figura 6. Análisis del atributo "NOMBRE_CURSO" con el filtro StringToNominal

Siguiendo con el orden de los atributos de tipo String tenemos el caso de "HORARIO_CURSO" del cual se obtiene la figura 7, en la que podemos apreciar lo siguiente: Los cinco horarios más utilizados para la programación de cursos en orden de mayor a menor frecuencia son: 07:00-10:00 con un número de frecuencia de diez y ocho, 11:00-14:00 con un número de frecuencia de diez y siete, 18:00-20:00 con un número de frecuencia de diez y seis, 07:00-09:00 y 08:00-11:00 con un número de frecuencia de quince, por último los horarios de 15:00-18:00, 18:00-21:00 y 11:00-13:00 tuvieron una frecuencia de trece. Los horarios que se programaron una sola vez en todo el Estado de Colima son: 13:18-15:33, 18:00-22:00, 09:00-11:30, 15:10-18:30, 19:00-21:00, 13:00-15:24 y 15:00-20:00.

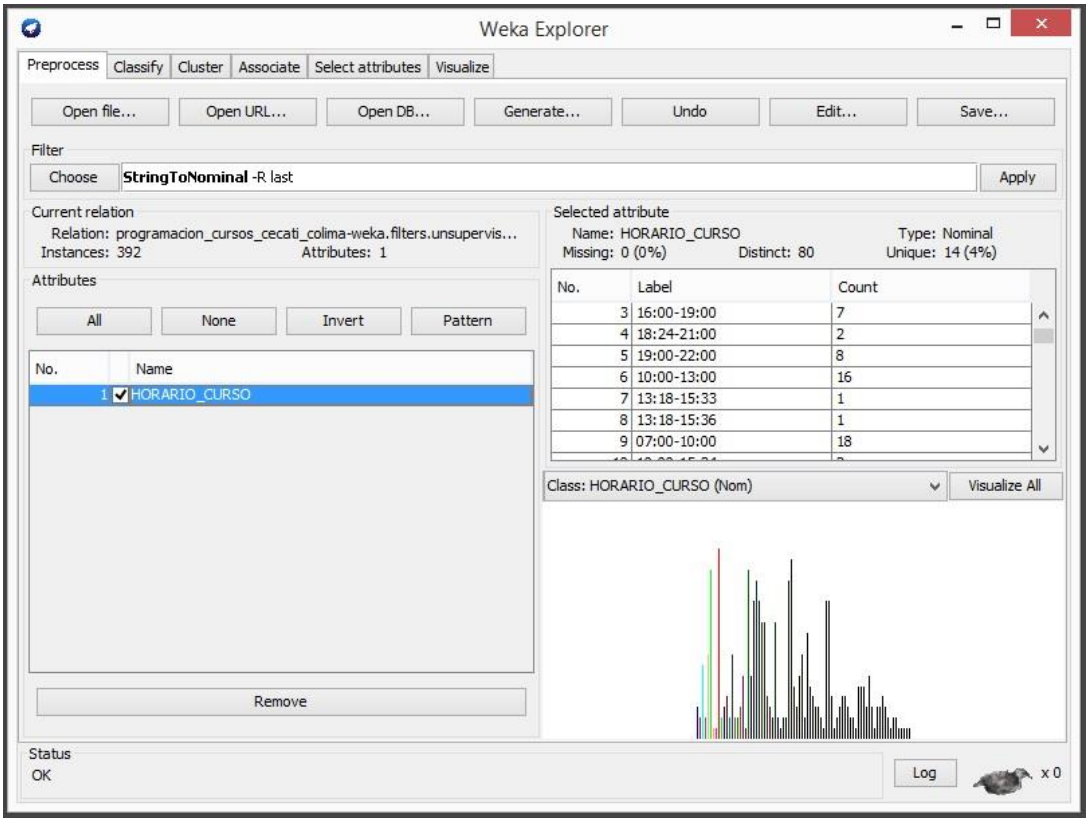

Figura 7. Análisis del atributo "HORARIO_CURSO" con el filtro StringToNominal

El atributo de "CECATI_No" por tener pocas instancias, es más fácil su análisis y no requiere de seleccionar un filtro o de remover los demás atributos, basta con que esté seleccionada la casilla correspondiente a "CECATI_No." y muestra la lista de instancias correspondientes, como se observa en la figura 8. Esta figura 8 muestra que el CECATI 145 fue el que programó más cursos durante el ciclo escolar que se está analizando con un total de 120 cursos lo que representa un 30.61 % del total de cursos programados en todo el Estado de Colima en ese ciclo escolar; en segundo lugar el CECATI 34 programó un total de 118 cursos en este ciclo escolar, lo que representó un 30.10% del total de cursos programados en todo el Estado de Colima; enseguida el CECATI 183 programó 97 cursos representando un 24.74% y por último el CECATI 126 con un total de 57 cursos y un 14.54% del total. Cabe hacer mención que estas cifras no corresponden a los cursos que se describen en la figura 6, porque ahí se mencionan diferentes nombres de cursos, mientras que aquí los nombres de los cursos pueden estar repetidos.

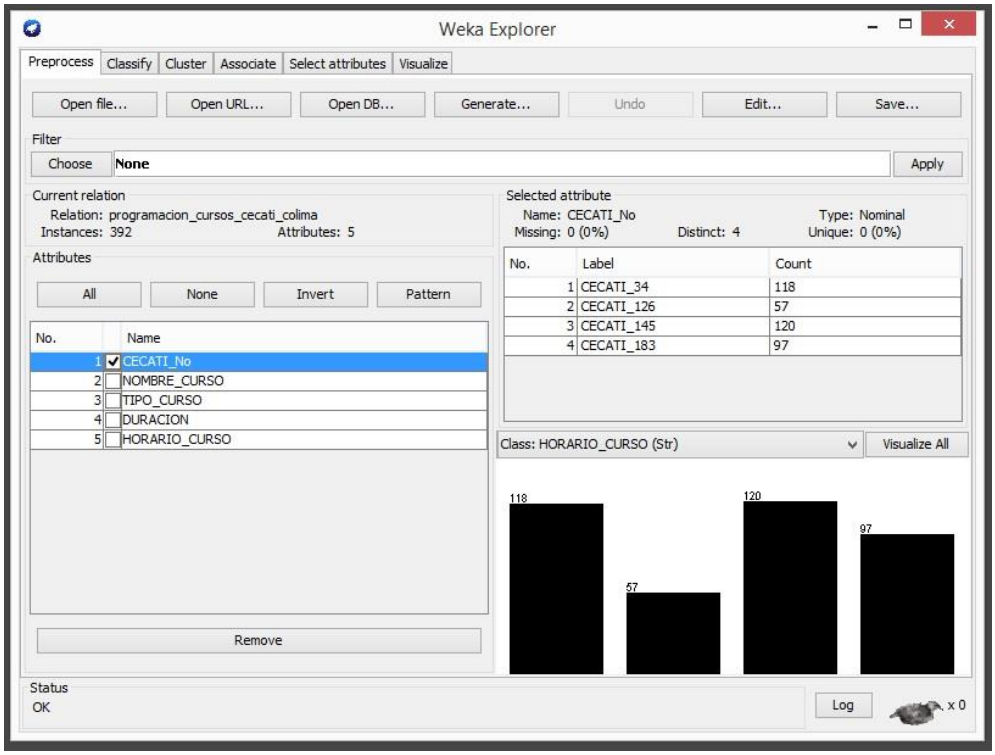

Figura 8. Análisis del atributo "CECATI_No."

El análisis del atributo "TIPO_CURSO" es similar al anterior, también tiene pocas instancias, como se aprecia en la figura 9 tiene cuatro instancias, a saber los cursos "REGULARES" hacen un total de 272 cursos programados de este tipo, lo que representa el 69.39% del total de cursos programados en el Estado de Colima durante el ciclo escolar 2013-2014; los cursos de "EXTENSIÓN" hacen un total de 97 cursos representando el 24.74% del total; los cursos "CAE" con un total de 19 cursos programados, representan un 4.85% en relación al total; por último los cursos programados como "ACCIONES_MOVILES" hacen un total de cuatro cursos, lo que representa el 1.02% del total de cursos programados en el Estado de Colima.

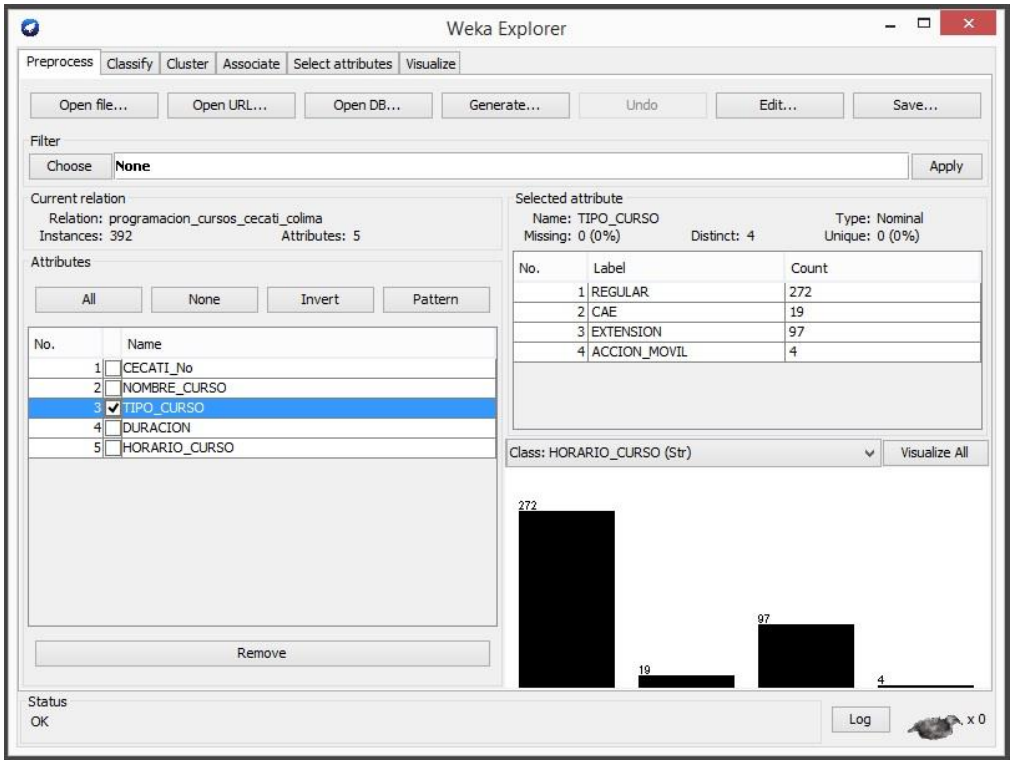

Figura 9. Análisis del atributo "TIPO_CURSO"

Para el caso del atributo "DURACION" (Figura 10) es conveniente utilizar el filtro "Discretize" para evitar que las instancias sean una lista enorme que no se pueda analizar, por lo tanto las cantidades aparecen por rangos, así los cursos con mayor duración en horas son en primer lugar los que se encuentran en el rango de 100 a 144 años, en segundo lugar el rango 56 a 100 horas de duración, y en tercer lugar el rango de 187 a 231 horas, el cuarto lugar el rango de 144 a 187 horas, el quinto lugar están los cursos con una duración de menos de 57 horas, el último lugar el rango de 362 a 406 horas de duración.

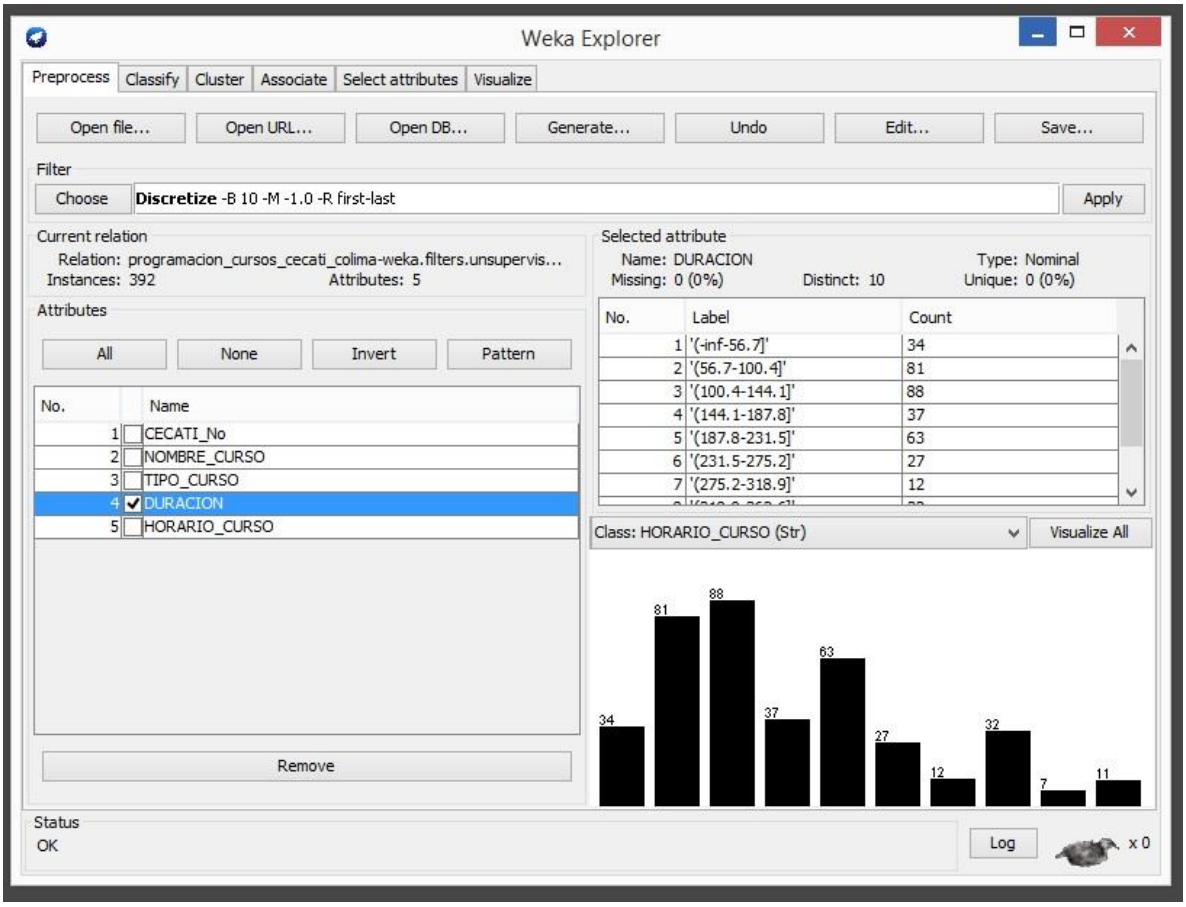

Figura 10. Análisis del atributo "DURACION"

Asociación de atributos

Para buscar algún tipo de asociación entre los atributos aplicamos la pestaña Associate de Weka y seleccionamos el filtro A priori y damos click en Start, si existe alguna asociación entre las instancias correspondientes a los cinco atributos, se despliega la lista de instancias que guardan cierta asociación, en este caso por ser independiente cada atributo no presentan ningún tipo de asociación como se muestra en la Figura 11.

Figura 11. Interfaz para Asociación de Atributos

Resultados

Los resultados presentados en el apartado del análisis de los datos representan datos confiables que quedan disponibles para su interpretación y la toma de decisiones por parte de los interesados. Estos datos tienen una base objetiva y fundamento científico y son mostrados tal y como los algoritmos de la aplicación Weka los presenta, sin sesgar los resultados, ni maquillarlos para provocar tendencias, los interesados serán los responsables de interpretarlos de manera adecuada para una toma de decisiones correcta.

Recomendaciones

Se recomienda usar la metodología empleada en este trabajo para realizar nuevos análisis de datos históricos correspondientes a otros ciclos escolares a nivel estatal. Asimismo, recomendamos ampliar la base de datos histórica para hacer análisis de datos históricos en un contexto nacional tomando en cuenta que los CECATI operan en toda la república mexicana.

Bibliografía

Dr. Sudhir B. Jagtap y Dr. Kodge B.G. "Census Data Mining and Data Analysis using Weka" (ICETSTM) International Conference in "Emerging Trends in Science, Tecnology and Management" 2013, Singapore.Recuperado de: http://arxiv.org/ftp/arxiv/papers/1310/1310.4647.pdf

Gonzalo Talo 25 de ene. De 2015. Copy of Tecnología y empresa. Recuperado de: https://prezi.com/6uhj2zx1izl3/copy-of-tecnologia-y-empresa/

Gonzalez Rojas, H.D.: "Importancia de la tecnología en las empresas" en Contribuciones a la Economía, febrero 2010. Recuperado de http://www.eumed.net/ce/2010a/

Dick Johnson 2011 "El Control de la Información: 6 empresas para gobernar a todos". Recuperado de: http://elespiritudeltiempo.org/blog/el-control-de-la-informacion-6-empresas-para-gobernarlos-a-todos/

D. Andrés Boza García, Dr. Angel Ortiz Bas, Dr. Eduardo VicénsSalort, Dña. Llanos Cuenca Gonzalez, "Data Warehouse para la gestión por procesos en el sistema productivo". Second World Conference on POM and 15tn Annual POM Conference, Cancun, Mexico, April 30 – May 3, 2004. Recuperado de: http://www.pomsmeetings.org/ConfProceedings/002/POMS_CD/Browse%20This%20CD/PAPERS/002-0278.pdf

Carmen de Pablos Heredero, Irene Albarrán Lozano, Guillermo Castilla Alcalá, "El Proceso de Implantación del Data Warehouse en la Organización: Análisis de un caso". Investigaciones Europeas de Dirección y Economía de la Empresa Vol. 4, 03,1998, pp. 73-92. Recuperado de: http://www.aedem-irtual.com/articulos/iedee/v04/043073.pdf

W.H. Inmon "Building the Data Warehouse: Getting Started". Recuperado de: http://www.academia.edu/3081161/Building_the_data_warehouse

Silberschatz, Abraham, Korth, Henry F. y Sudarshan, S. "Fundamentos de Bases de Datos". McGrawHill, 4a Ed., 2002.

I.Olmos Pineda y J.A. González Bernal "Casos de éxito de Minería de Datos". Recuperado de: http://es.scribd.com/doc/93421745/Caso-de-Exito-Mineria-de-Datos#scribd

Webmining Consultores "KDD: Proceso de Extracción de conocimiento". 10 de Enero de 2011 • En la Categoría Business Intelligence & Analytics, Data Mining. Recuperado de: http://www.webmining.cl/2011/01/proceso-de-extraccion-de-conocimiento/

José Manuel Molina López y Jesús García Herrera "Técnicas de Analisis de Datos Aplicaciones Prácticas utilizando Microsoft Excel y Weka" 2006. Recuperado de: http://www.giaa.inf.uc3m.es/docencia/II/ADatos/apuntesAD.pdf

Diego García Morate "Manual de Weka" licencia CreativeCommons Reconocimiento-NoComercial-SinObraDerivada2.0 .Recuperado de: http://www.metamotion.com/diego.garcia.morate/download/weka.pdf