

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

Análise de Dados Históricos Curso de Programação em Colima Estado CECATI

Manuel Espinosa Ortega

Instituto Tecnológico de Colima, México

g7946192@itcolima.edu.mx

Nicandro Farías Mendoza

Instituto Tecnológico de Colima, México

nmendoza@ucol.mx

Jesús Alberto Verduzco Ramírez

Instituto Tecnológico de Colima, México

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.

Palabras clave: Minería de datos, Data Warehouse, datos históricos, proceso KDD.

Resumo

Hoje, as ferramentas para a extração de informações estão a melhorar o processo para as empresas e agências podem obter informações a partir de grandes volumes de dados. Sistemas de extração de informações é tradicionalmente aplicado como uma seqüência de módulos de extração para fins especiais torna-se, como uma classe particular de peças de informação relevantes que são usados por agências ou empresas a tomar decisões que melhoram a processos funcionalidade. Este documento fornece uma visão geral do Sistema

Web para cursos de programação em CECATI (SWPCC) é feita. Em particular, a nossa investigação centra-se no desenvolvimento de um módulo para extração de informações a partir da análise de dados históricos, agendar cursos CECATI no Estado de Colima, durante o ano escolar 2013-2014 por ferramentas de Data Warehouse e Data Mining. A análise dos dados históricos produz informações sobre os cursos regulares, mal planejado, e outras áreas de oportunidade e de outros aspectos como o tempo ea duração dos cursos que podem influenciar a demanda por cursos ministrados em CECATI , o que nos permite tomar decisões para cursos de planejamento eficazes. A metodologia utilizada para a nossa investigação, consiste nas seguintes três fases:

Na primeira fase, a contextualização desta pesquisa é apresentada, descrevendo os mecanismos e técnicas de extração de informações, a exposição dos bancos de dados históricos, System Web para cursos de programação em CECATI (SWPCC) e módulo extracção de informação.

Em uma segunda fase apresenta em detalhe cada uma das etapas que foram executadas para obter um conjunto de dados limpo que poderiam ser analisados por uma ferramenta chamada Weka (Waikato Ambiente para a análise do conhecimento) (Sudhir, Kodge, 2013) a partir de um origem conjunto de dados de CECATI no Estado de Colima o ano letivo 2013-2014, a conversão do formato de arquivo original, processamento de dados, análise de dados, filtragem de trabalho e discretização os dados.

Numa terceira fase detalhada cada um dos resultados obtidos na análise de dados utilizando diferentes algoritmos que tem Weka, buscando uma apresentação de forma simples e clara a estes resultados, de modo que eles podem fornecer o novas informações interessado para a tomada de decisão. A metodologia que é detalhado neste documento pode servir como uma base para pesquisas futuras com outros ciclos escolares, com o objectivo de obter novos conhecimentos.

Palavras-chave: Mineração de Dados, Data Warehouse, dados históricos, o processo de KDD

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

Introdução

Muitas empresas não dão a importância que se deve à utilização de tecnologia por isso é importante para o desenvolvimento do mesmo, como é hoje exige que as empresas para competir no mercado electrónico e dispõem de sistemas de informação adequada suas necessidades (Talo, 2015). Hoje as empresas precisam ter (SI) sistemas de informação que lhes permitam obter informação fiável e ajudá-los a tomada de decisão (Rojas, 2010), de fato, há a alegação de que quem controla a informação vai dominar o mundo (Johnson, 2011), e é evidente que, actualmente, existem grandes quantidades de informação com esse fundamento para controlar esses grandes volumes de informações necessárias uso de ferramentas para ser capaz de extrair, classificar, triagem, limpeza, estrutura, processo, etc. . de modo que esta informação torna-se útil para a tomada de decisão. Por esta necessidade surgiu mecanismos como armazém de dados e mineração de dados que permitem que essas funções em nome das instituições.

Data Warehouse

Depois das dificuldades dos sistemas tradicionais para atender às necessidades de informação, o conceito de Data Warehouse surge como uma solução para as necessidades informacionais globais da empresa (Boza, 2004 et al). Data Warehouse O termo foi cunhado pela primeira vez por Inmon (Pablos, Albarran, Castilla, 1998), traduz literalmente como Data Warehouse. No entanto, se o data warehouse foi apenas um armazenamento de dados, problemas permaneceria o mesmo como nos Centros de Informação¹. A principal vantagem destes sistemas é baseado no conceito fundamental, a estrutura de informação. Este conceito significa o armazenamento de informação homogénea e confiável em uma estrutura baseada na consulta e tratamento aninhada na mesma, e em um ambiente diferenciado de sistemas operacionais. Conforme definido Inmon (2000), o Data Warehouse é caracterizado por:

¹centros especializados, criados com a finalidade de coletar dados, produzir informações e disponibilizá-lo para todas as instituições, universidades, sindicatos e associações empresariais, bem como para a cooperação internacional. Rosanna Silva "Informação e Documentação Centers (México) www.monografias.com

- **Integrada:** dados armazenados no data warehouse devem ser integrados em uma estrutura consistente, de modo que as inconsistências existentes entre os vários sistemas operacionais devem ser eliminadas. A informação é muitas vezes também estruturada em diferentes níveis de detalhe para atender diferentes necessidades dos utilizadores.
- **Tema:** apenas os dados necessários para o processo de geração de conhecimento do negócio estão integrados no ambiente operacional. Os dados são organizados por tópico para facilitar o acesso e compreensão pelos utilizadores finais. Por exemplo, todos os dados do cliente podem ser consolidadas em uma única tabela do Data Warehouse. Deste modo, os pedidos de informação do cliente será mais fácil para responder porque toda a informação reside no mesmo local.
- **Histórico:** o tempo é uma parte implícita das informações contidas em um Data Warehouse. Em sistemas operacionais, os dados reflectem sempre o estado da atividade empresarial no presente. Pelo contrário, as informações armazenadas no Data Warehouse serve, entre outras coisas, para a análise de tendências. Portanto, o armazém de dados é carregado com diferentes valores de uma variável no tempo para permitir comparações.
- **Não voláteis:** armazenam informações em um data warehouse existe para ser lido, não modificada. A informação é, portanto, permanente, ou seja, a atualização do data warehouse incorporando os valores mais recentes, tendo as variáveis contidas nele sem qualquer ação sobre o que já existia.

Minería de Datos

Silberschatz, Abraham, Korth, Henry F. y Sudarshan, S. (2002) definir Data Mining (DM) como um processo automático ou semi-automático que busca descobrir padrões escondidos em um conjunto de dados e também potencialmente útil para os usuários do banco de dados (BD). Nos MD contempladas diversas estratégias para identificar os diferentes tipos de padrões, tais como árvores de classificação, redes neurais, redes Bayesianas, técnicas de associação, entre outros (Olmos Gonzalez, 2007). O objetivo em todo o processo PI é a obtenção de padrões de interesse para o usuário final. Para conseguir

isso, é necessário preparar adequadamente os dados para processamento, escolha de um método adequado para extrair os padrões desejados e finalmente determinar a forma de avaliar os padrões encontrados.

Extração de conhecimento está associada principalmente com o processo de descoberta conhecido como Knowledge Discovery in Databases (KDD). É um processo que extrai a informação da qualidade que pode ser usado para tirar conclusões a partir das relações ou padrões dentro dos dados (webmining, 2011). A figura a seguir ilustra as etapas do processo de DCBD (Figura 1):

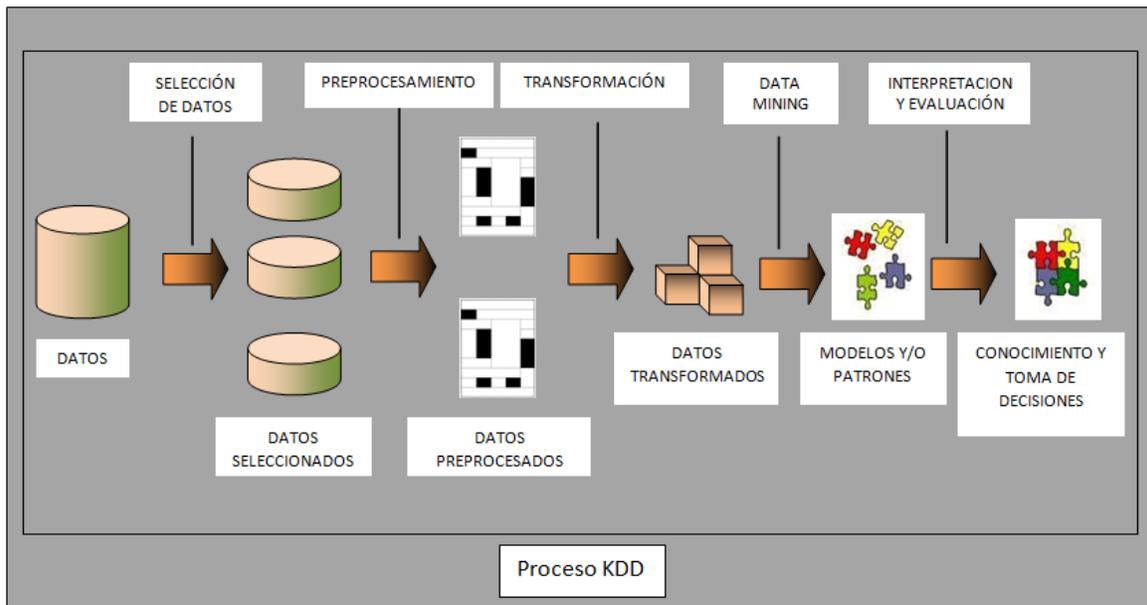


Figura 1. Etapas del proceso KDD

Como mostrado na figura acima, os passos do processo de DCBD é dividido em cinco fases são:

Seleção de Dados

Nesta fase, as fontes de dados e os tipos de informações a serem utilizadas são determinados. É a fase em que os dados relevantes para a análise são extraídos das fontes de dados ou.

I pré-processamento

Este passo é a preparação e limpeza de dados extraídos de diferentes fontes de dados de uma forma exequível, necessários para as fases posteriores. Nesta fase várias estratégias são

usados para manipular dados em falta ou em branco, ou dados inconsistentes que estão fora de gama, resultando em a extremidade de uma estrutura de dados adequada para processamento adicional.

Transformação

Ele consiste no processamento preliminar dos dados, processamento e geração de novas variáveis de estrutura de dados existente com o caso. Aqui operações de agregação ou de normalização são executadas, a consolidação de dados a partir de uma forma necessária para a próxima fase.

Data Mining

É a fase de modelagem em si, onde os métodos inteligentes são aplicadas com o objectivo de extrair até então desconhecida, válido, romance, potencialmente úteis e compreensíveis padrões que estão contidas ou "ocultas" nos dados.

Desempenho e Avaliação

Os padrões obtidos são identificados e são muito interessante, com base em algumas medições e a avaliação dos resultados é executada.

Bases de dados históricas

Centros de Formação para o trabalho industrial (CECATI) têm um sistema chamado "Curso de Programação do Sistema Web em CECATI" (SWPCC) .Com este sistema (Figura 2) é executada automaticamente agendamento de cursos CECATI ensinado em cada ano escolar.

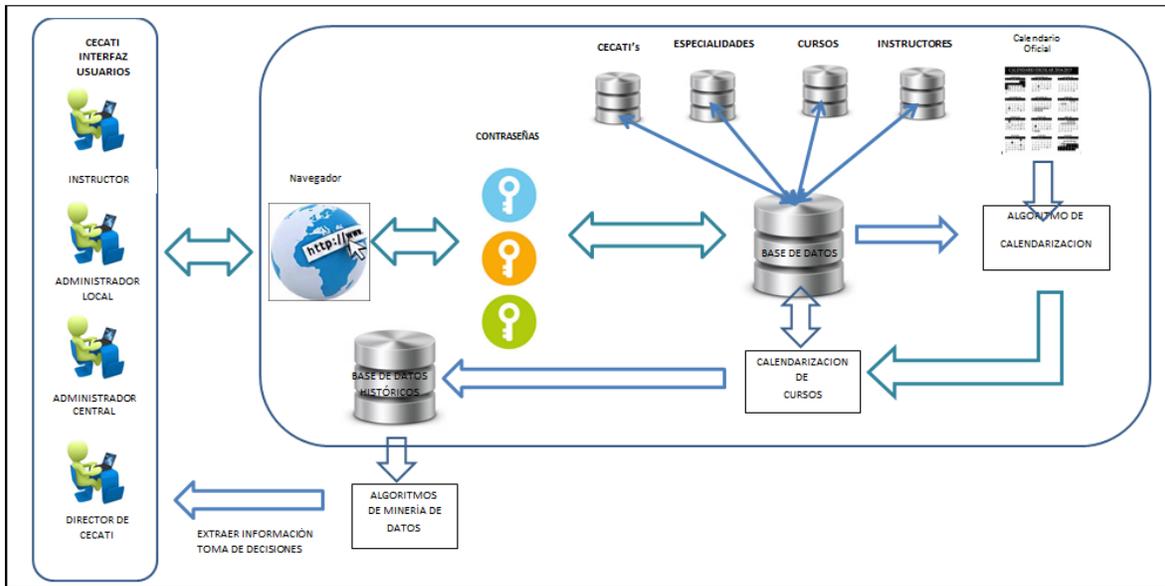


Figura 2. Sistema Web para la Programación de Cursos en los CECATI

O sistema também possui um módulo de extração de informação compreende os seguintes componentes: Em primeiro lugar, a base de dados histórica de planeamentos do curso. Esta base de dados é gerado com o poder de cursos de programação, que são oferecidos em todo o país por ano-a-ano em cada CECATI. Em segundo lugar, um processo de aplicação de algoritmos para extrair informações e em terceiro lugar, uma interface para a interpretação e visualização da informação extraída (Figura 3).

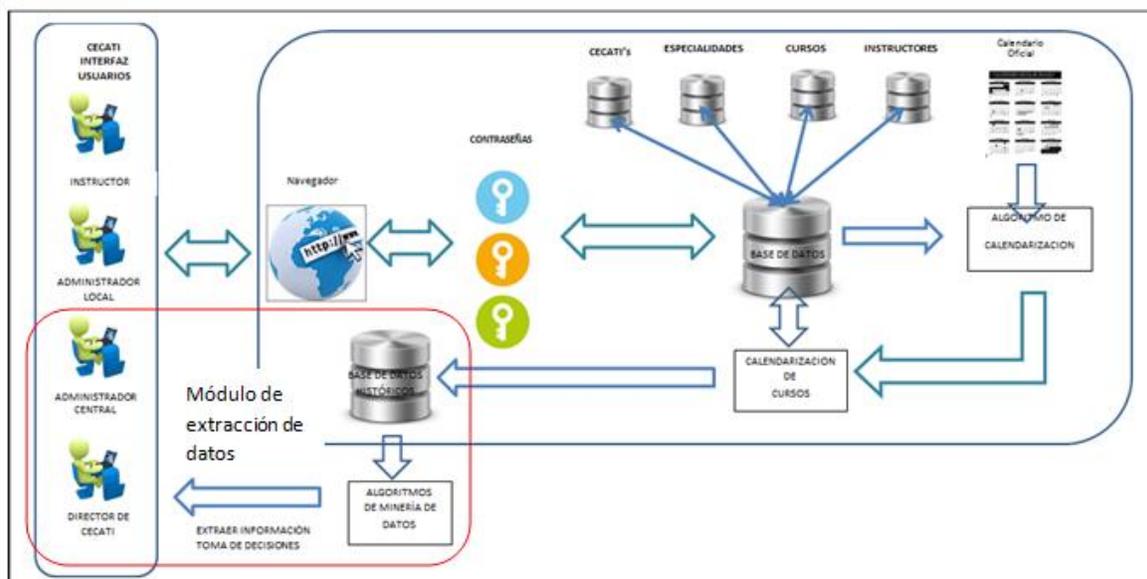


Figura 3. Módulo de extracción de información

Esta base histórica para analisar as informações utilizando ferramentas de Data Warehouse e Data Mining para fins de obtenção de padrões, tendências ou projeções estatísticas para ajudar (Molina, García, 2006) de tomada de decisão. Nesta pesquisa Weka foi escolhido para ser um pedido de licença livre, resultados muito confiáveis e porque tem a maioria dos algoritmos utilizados (árvores de classificação, redes neurais, redes Bayesianas, técnicas de associação, etc.) para o a análise de dados. Para realizar a análise dos dados aqui mencionados, são levados em conta apenas as despesas efectuadas no Estado de Colima CECATI para facilitar o acesso a essas informações.

Obtendo os fatos

O banco de dados histórico do sistema SWPCC é a fonte para a obtenção dos dados de origem para analisar por Weka. O formato "xls" é para arquivos que podem ser abertos com o Excel, os dados de origem são formato ".sql", você deve importar o programa Excel para alterar o formato original ".sql" formato ".xls" . A razão para isto é que a conversão do Excel torna a limpeza dos dados, deixando apenas os dados que são úteis para a análise a ser executada usando Weka.

Limpar a dados

Através do Excel campos de nome de reposição ferramenta para todos os atributos que têm dados escritos de forma diferente unificada. Além de unificar o conteúdo dos dados, atributos que não são necessários para a análise do conjunto de dados, por exemplo eliminado: o atributo de "endereço", "colónias", "postal_code", "aparelho", "NOMBRE_JEFE_AREA" seu conteúdo são desnecessários para a informação pretendida a ser obtida; o atributo "especialidade" é removida porque você pode adivinhar o nome dos cursos programados; também "dias não-úteis" também removida porque seu conteúdo não é relevante.

Discretização dados

O conjunto de dados a serem analisados contém alguns campos que são numérico, por exemplo, "DURACION_EN_HORAS"; para o processo de análise com Weka são obrigados a ter um carácter simbólico, pois, necessário discretização deste tipo de dados, antes que esta ação deve primeiro converter o aparelho para o tipo de formato delimitado

por vírgulas CSV dados limpos, em seguida, copiar o atributos e tipo de arquivo de dados .arff (Diego Garcia) a seguinte estrutura (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

No top comentários relevantes precedidos pelo sinal% é colocada. Em seguida, o nome da relação precedido por "@ relacionamento" está escrito. Posteriormente enumerados sob forma de coluna de cada um dos atributos precedidas por "atributo", seguido do tipo de dados: real, inteiro, ou corda, conforme o caso. Esta estrutura é reconhecido por WEKA e deve ser escrupuloso na sintaxe de cada conceito como qualquer mudança em si não permitem que você abra o arquivo de analisar. Underdata percorrer todo delimitado por vírgulas dados, como pode ser visto no exemplo que se segue (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

Cuidar dessa estrutura ea sintaxe correta, o programa Weka (que é muito escrupulosa) recebe a informação sem restrição permitindo utilizar todas as funções disponíveis para analisar o conjunto de dados.

A análise dos dados

O conjunto curso de programação em CECATI no Estado de Colima neste documento dados serão analisados pela ferramenta Weka ele contém os seguintes atributos: CECATI_No, NOMBRE_CURSO, TIPO_CURSO, duração, HORARIO_CURSO, com o qual é pedida informação relevante sobre os cursos programados em CECATI no Estado de Colima durante o ano 2013-2014 escola. Para a análise do conjunto de dados algumas orientações são seguidas: Os atributos cujo conteúdo é uma lista de opções e também muito grande exigindo que contém mais de uma palavra, eles devem ser do tipo String como o caso do "NOMBRE_CURSO" deve ser removido outros atributos e deixar apenas o atributo a ser analisado, em seguida, selecione o StringToNominal Weka pode filtrar para exibir a lista de instâncias que contém o atributo "NOMBRE_CURSO". Tal é o caso do atributo "HORARIO_CURSO", que é do tipo String, assim, para a análise é igualmente

aplicável. Em seguida, os números de cada um destes atributos, e os resultados obtidos mediante a aplicação do filtro mostrado StringToNominal:

O valor resultante (Figura 6) mostra um total de 111 cursos diferentes agendada durante o ano 2013-2014 escola em CECATI no Estado de Colima, dos quais cursos serviços especializados de beleza e corte de cabelo styling e maquiagem rosto CORES E TRANSFORMAÇÃO no cabelo, são o que está programado mais frequentemente em relação a outros cursos de outras especialidades. Depois neste campo, especializando-se em refrigeração e A / C com manutenção do sistema A / C e campo de refrigeração, mostra segundo programa frequência deste curso, a especialidade de electricidade está em terceiro lugar com o curso chamado INSTALAÇÃO RESIDENCIAL ELÉTRICA. Entre os cursos regulares que foram programadas apenas uma vez em todo o estado de Colima nesses centros de treinamento incluem: poluentes de emissões de motores elétricos de controle de manutenção, sistema de reparo CLUTCH sistema de reparo, reparo do freio ABS, mantendo uma redes de área local (LAN), de preparação de bebidas e de serviços para os hóspedes.

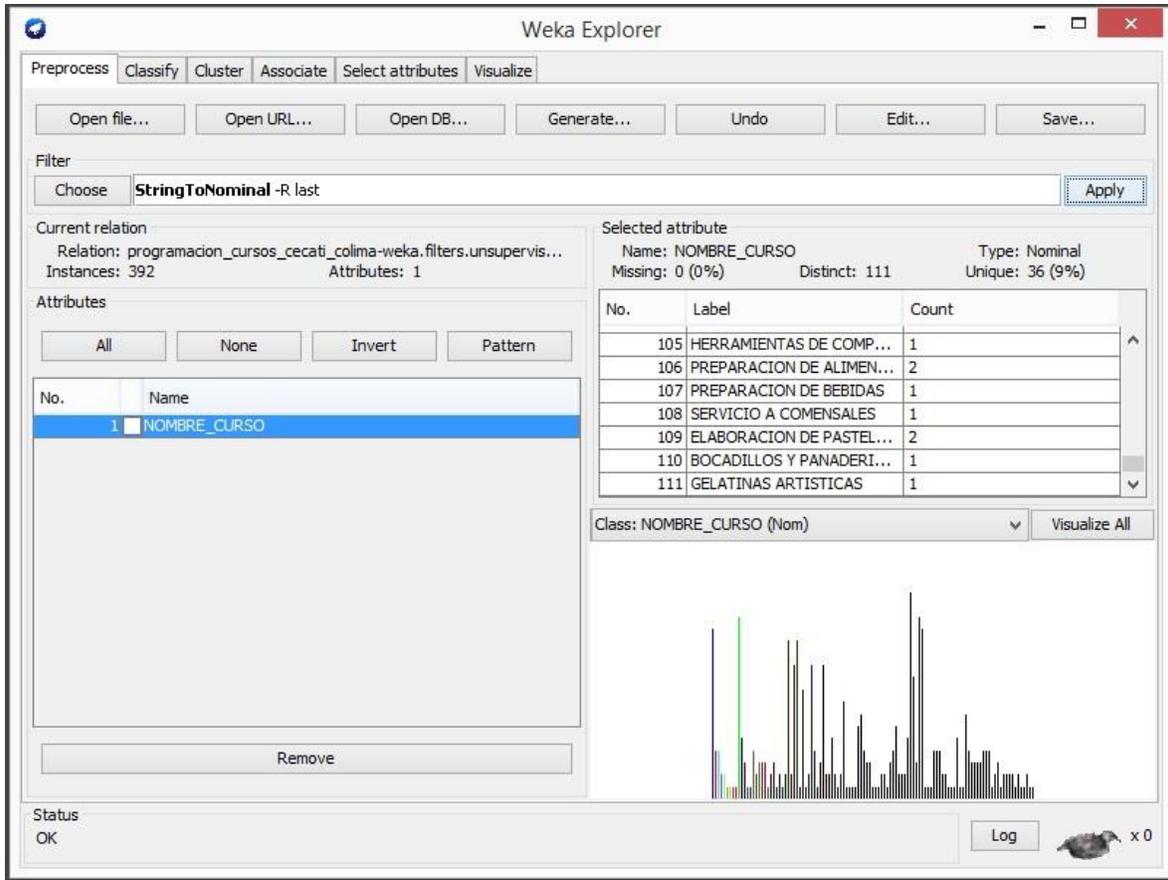


Figura 6. Análisis del atributo "NOMBRE_CURSO" con el filtro StringToNominal

Seguindo a ordem dos atributos do tipo String, temos o caso de "HORARIO_CURSO", que figura 7, podemos ver que o seguinte é obtido: os cinco mais usados para cursos de programação em ordem da maior para a menor frequência horários eles são: 07: 00-10: 00, com um número de frequências de dezoito, 11: 00-14: 00, com um número de frequência dezassete, 18: 00-20: 00, com um número de todos os dez seis, 07: 00-09: 00 e 08: 00-11: 00 com um número de frequência de quinze vezes, finalmente, 15: 00-18: 00, 18: 00-21: 00 e 11:00 -13: 00 tinham uma taxa de treze. Os tempos foram programadas apenas uma vez em todo o estado de Colima são: 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.

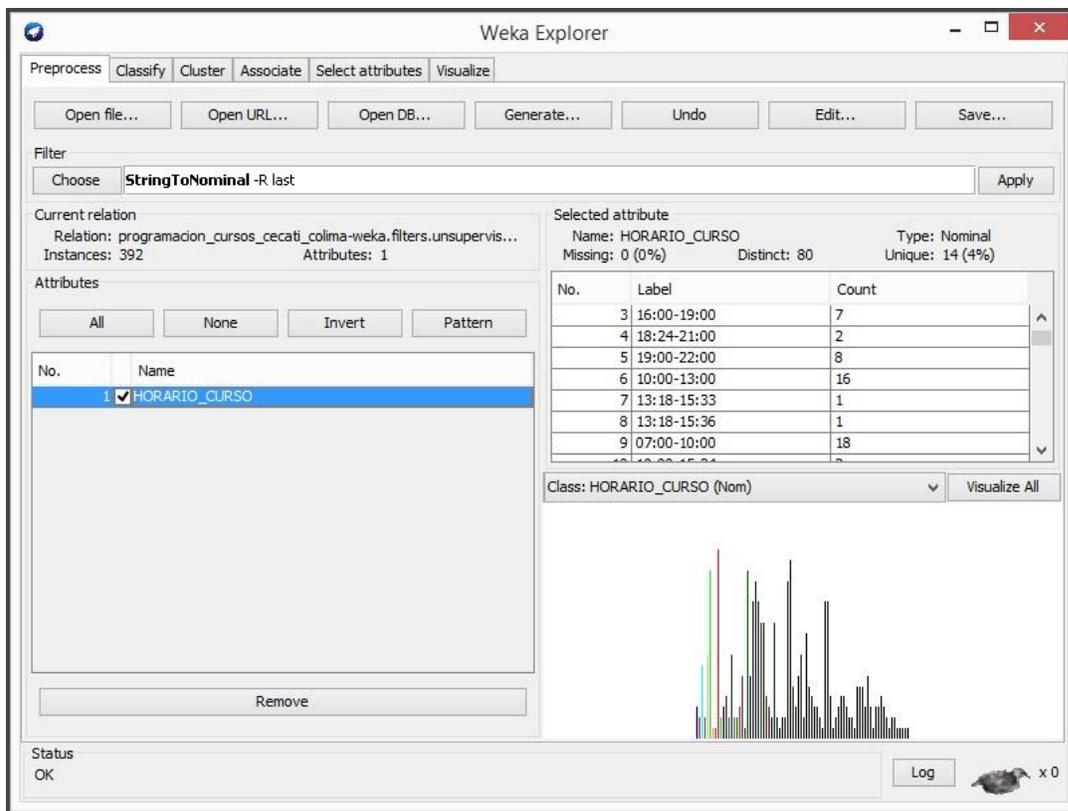


Figura 7. Análisis del atributo "HORARIO_CURSO" con el filtro StringToNominal

O atributo "CECATI Não" ter poucos casos, a sua análise é mais fácil e não requer a seleção de um filtro ou remover outros atributos, apenas certifique-se a caixa está marcada para "CECATI_No." E mostra a lista de instâncias correspondente, como mostrado na Figura 8. Esta Figura 8 mostra que o 145 era o CECATI mais cursos programados durante o ano escolar sendo analisados um total de 120 cursos representando um 30.61% do total programada no todo o Estado de Colima nesse ano lectivo; em segundo lugar a 34 CECATI agendado um total de 118 cursos este ano escolar, o que representou 30,10% de todos os cursos programados em todo o Estado de Colima; uma vez que os programados 97 cursos CECATI 183, representando 24,74% e, finalmente, o CECATI 126 com um total de 57 cursos e 14,54% do total. Deve mencionar-se que estes valores não correspondem aos campos descritos na Figura 6, porque o referido curso nomes diferentes são mencionados, enquanto que aqui os nomes dos campos pode ser repetido.

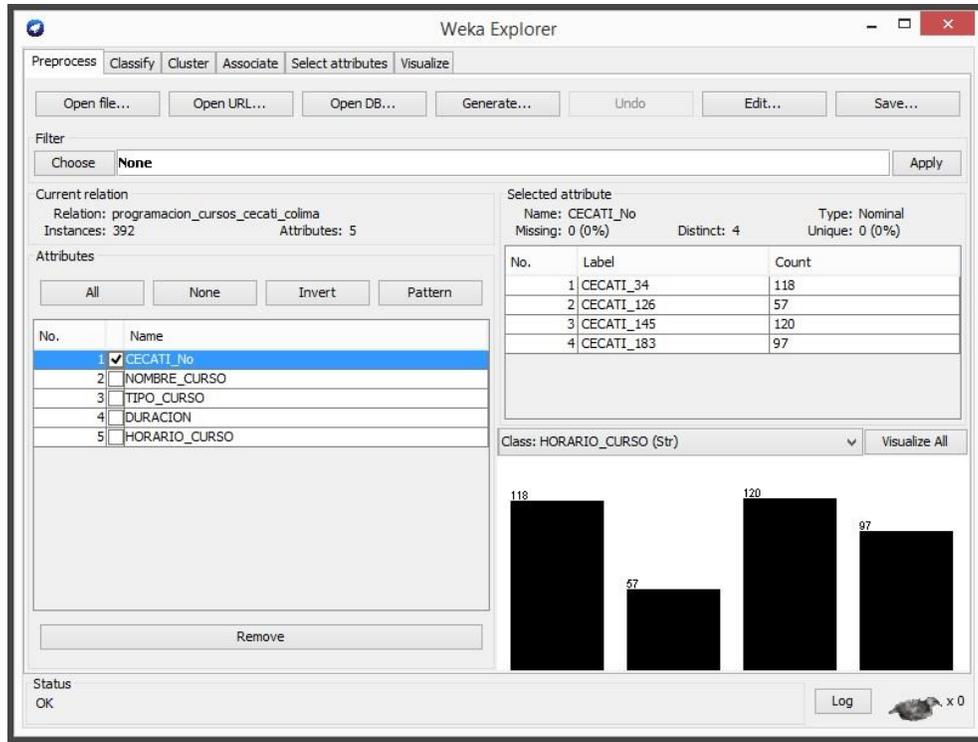


Figura 8. Análisis del atributo "CECATI_No."

A análise atributo "TIPO_CURSO" é semelhante à acima, também tem alguns casos, tal como mostrado na Figura 9 tem quatro instâncias, nomeadamente cursos

"Regular" para um total de 272 cursos regulares deste tipo, representando 69,39% de todos os cursos programados no Estado de Colima durante o ano 2013-2014 escolar; cursos de "estimar" para um total de 97 cursos que representam 24,74% do total; cursos "CAE" com um total de 19 cursos programados representam um 4,85% em relação ao total; cursos finalmente agendada como "AÇÃO MOBILE" perfazendo um total de quatro cursos, o que representa 1,02% dos cursos regulares no estado de Colima. No caso de "duração" atributo (Figura 10), é conveniente usar-se o filtro

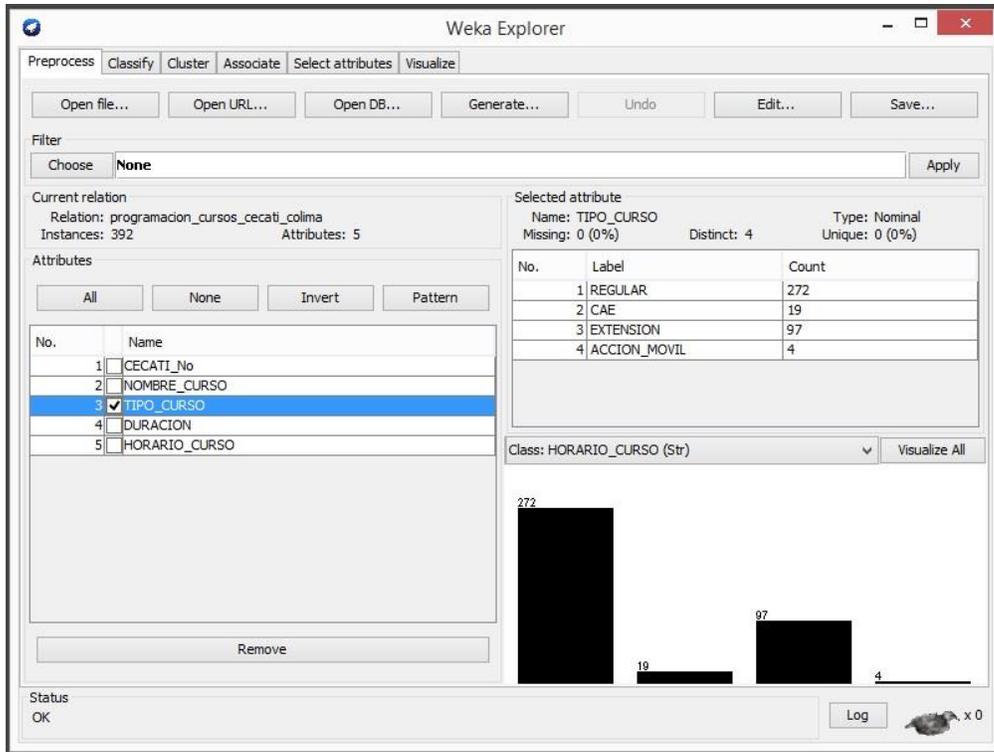


Figura 9. Análisis del atributo "TIPO_CURSO"

"Discretizar" prevenir os casos são uma lista grande que não pode ser analisada, assim, as quantidades e os intervalos listados cursos com mais horas são em primeiro lugar aqueles encontrados na gama de 100-144 anos, segundo faixa de 56-100 horas, e em terceiro lugar a faixa de 187-231 horas, a quarta faixa de 144-187 horas, o quinto lugar são os cursos com uma duração inferior a 57 horas, o último lugar da faixa de 362-406 horas.

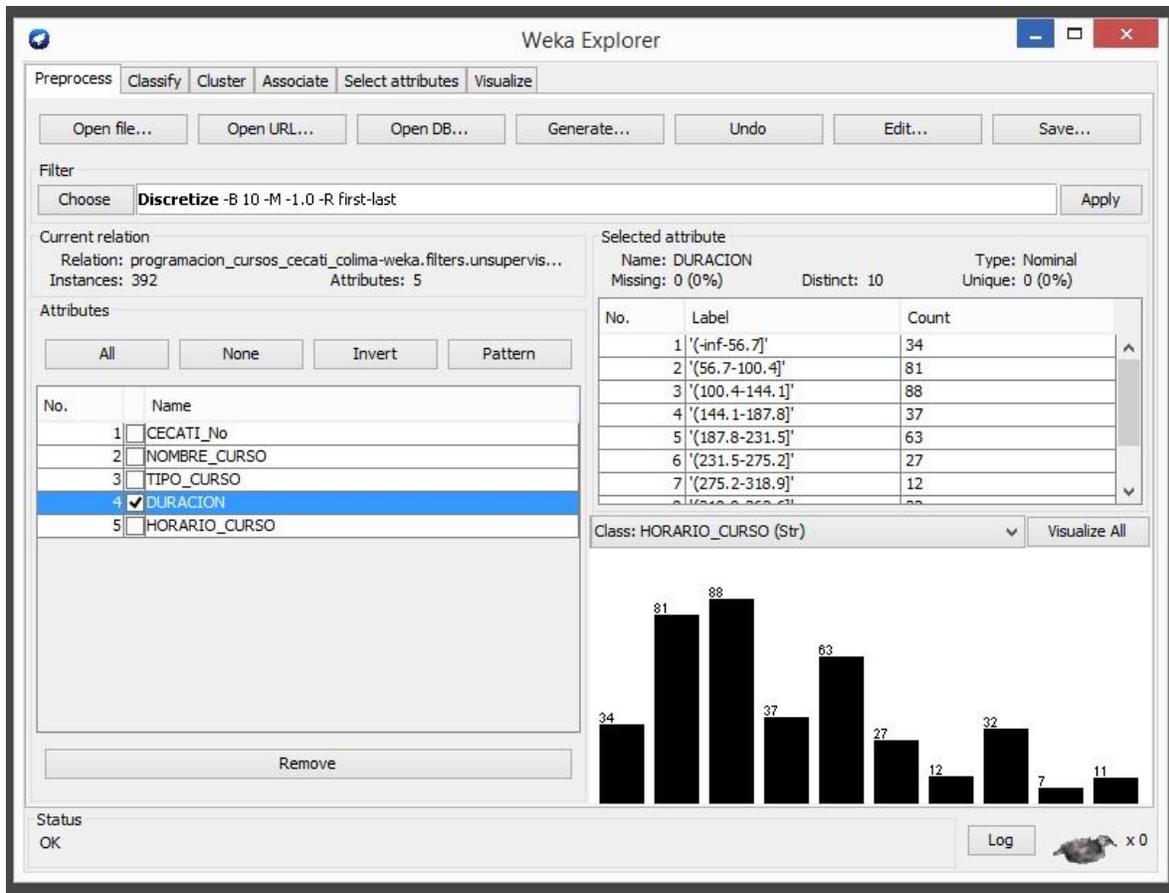


Figura 10. Análisis del atributo “DURACION”

Atributos de associação

Para procurar algum tipo de associação entre os atributos se aplicam a guia Associate Weka e seleccione o filtro A priori e nós clique em Iniciar, se há alguma associação entre as instâncias correspondentes aos cinco atributos, a lista de instâncias que têm alguma associação se desenrola neste caso, para ser independente de cada atributo não exibem qualquer associação, tal como mostrado na Figura 11.

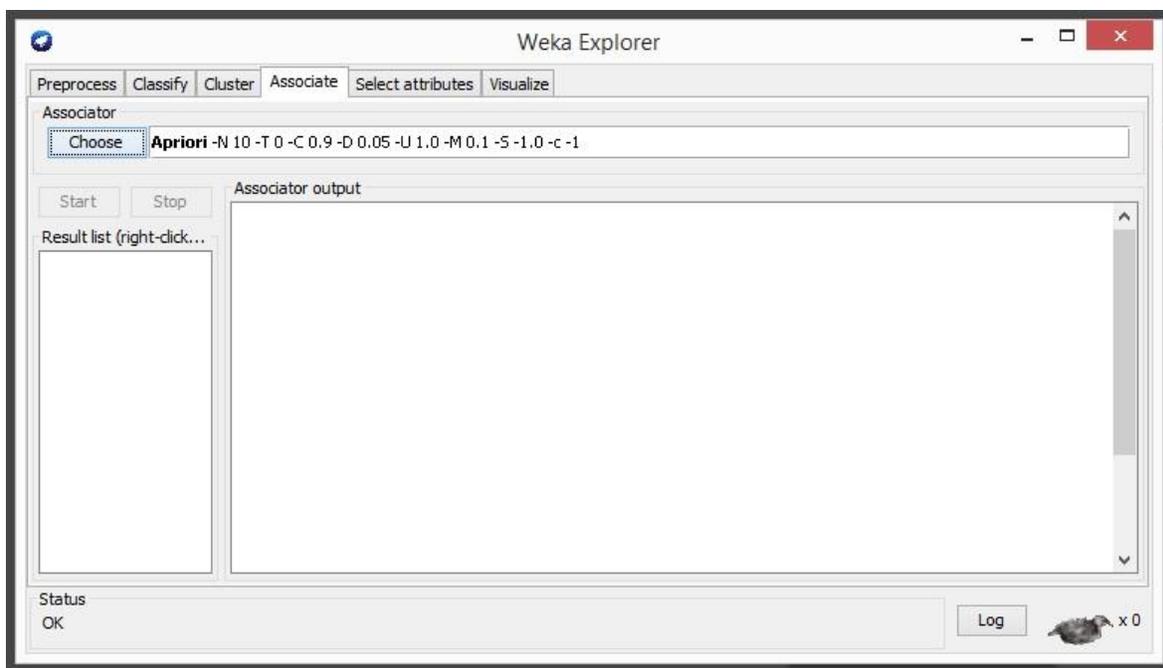


Figura 11. Interfaz para Asociación de Atributos

Resultados

Os resultados apresentados na seção sobre análise de dados representam dados confiáveis estão disponíveis para interpretação e tomada de decisões pelas partes interessadas. Esses dados têm uma base objetiva e fundamento científico e são mostrados como algoritmos de Weka aplicar apresentadas sem viés os resultados, nem maquillarlos a causar tendências, as partes interessadas serão responsáveis por interpretá-los corretamente para uma decisão correta.

Recomendado

Recomendamos a utilização da metodologia utilizada neste trabalho para análise posterior para outros anos escolares em todo o estado de dados históricos. Recomendamos também a expansão da base de dados históricos para análise de dados históricos, num contexto nacional, tendo em conta que o CECATI operando em todo o México.

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, Technology 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 15th 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>