

Modelo de Clustering basado en Redes Neuronales para identificar el perfil de los alumnos por segmento enfocado a los servicios de Tecnologías de Información de la Universidad Peruana Unión

Moreno Tineo, Ronald Jesus; Pisco Sandoval, Edward Ajax; Vela Becerra, Cristian Irvin
Dr. Palza Vargas, Edgardo; Dr. Mamani Apaza, Guillermo; Mg. Acuña, Erika

Resumen

En la presente investigación se ha aplicado el algoritmo de k-medias del modelo de redes neuronales artificiales, para agrupar e identificar las características que perciben los alumnos con respecto a los servicios de tecnologías de información de la UPeU.

El algoritmo ha determinado 3 grupos, y por cada grupo ha identificado los atributos y características más relevantes de los servicios que brinda DIGESI. Para lograr el objetivo se utilizó la metodología CRISP-DM que es para proyectos de minería de datos.

CRISP-DM como metodología muy difundida y con los antecedentes de trabajo anteriores, permite un mayor respaldo al desarrollo del proyecto.

Palabras clave: Algoritmo de K-medias, perfil de alumno, redes neuronales, minería de datos, tecnologías de información, metodología CRIP-DM.

Model-based Clustering Neural Networks to identify the profile of students by focusing on the services segment of Information Technology of **Universidad Peruana Unión**

Abstract

In this research we applied the k-means algorithm of artificial neural network model, to group and identify the characteristics that students receive with respect to information technology services of the UPeU. The algorithm has identified 3 groups and each group has identified the most relevant attributes and characteristics of the services provided DIGESI. To achieve the objective, the methodology CRISP-DM is for data mining projects. Methodology CRISP-DM as widespread and previous work history, allows greater support for the project.

Keywords: Classification trees, credit risk.

Introducción

El propósito de la investigación es agrupar por atributos comunes la percepción de los servicios de tecnologías de información aplicando los fundamentos de redes neuronales artificiales, en particular el algoritmo de k-medias.

Para desarrollar esta investigación se utilizó la metodología de minería de datos CRISP-DM que consta de 6 fases y son: entender el negocio, entender los datos, preparar los datos, modelar,

evaluar y desplegar (Ref-01); la herramienta informática que se utilizó para validar el modelo ha sido el SPSS Clementine, la cual contiene algoritmos de redes neuronales artificiales supervisadas y no supervisadas.

En la actualidad, la Universidad Peruana Unión, brinda servicios informáticos dentro del campus, y estos servicios son: internet, portal académico, laboratorios de cómputo entre otros.

Dado que la universidad desea mejorar los

servicios de TI (Tecnologías de Información) brindados, es imprescindible conocer las deficiencias que existen, y las demandas de los alumnos, los cuales necesitan ser segmentados

DIGESI (Dirección General de Sistemas), no conoce cuáles son las características o atributos que tienen mayor relevancia ante la percepción del alumno respecto a los servicios de TI, debido a la existencia de diferentes carreras profesionales (educación, salud, ingeniería etc). La importancia y la necesidad de los alumnos pertenecientes a una carrera varía; por ejemplo, un estudiante de ingeniería de sistemas le daría más importancia a la variedad de programas instalados en los laboratorios de cómputo frente a un alumno de salud. Es ahí donde nace la necesidad de identificar a grupos de alumnos con necesidades comunes para así ofrecer un servicio de TI personalizado.

El presente trabajo de investigación tiene como objetivo identificar el perfil de los alumnos respecto a los servicios aplicando un algoritmo de clústeres basado en red neuronal k-media para la segmentación de los alumnos de la UPeU, la segmentación de los alumnos en grupos con comportamientos parecidos.

Antecedentes de Estudio

Uso de minería web para identificar perfiles de usuarios

Se realizaron investigaciones con minería de datos para reconocer patrones y perfiles de usuarios, para esto se utilizó la lógica difusa como herramienta principal y se centró en el análisis del comportamiento de los usuarios en la web, para poder obtener una representación de este perfil (Escobar 2007).

La British Broadcasting Corporation (BBC) del Reino Unido emplea un sistema para predecir el tamaño de las audiencias televisivas para un programa propuesto, así como el tiempo óptimo de exhibición (Brachman y otros, 1996). El sistema utiliza redes neuronales y árboles de decisión aplicados a datos históricos de la cadena para determinar los criterios que

participan según el programa que hay que presentar. La versión final se desempeña tan bien como un sistema experto con la ventaja de que se adapta fácilmente a los cambios porque es constantemente reentrenada con datos actuales.

En universidades - Los titulados llevan a cabo actividades profesionales relacionadas con sus estudios

Se hizo un estudio sobre los recién titulados de la carrera de Ingeniería en Sistemas Computacionales del Instituto Tecnológico de Chihuahua II, en México (Rodas, 2001). El objetivo era saber si con los planes de estudio de la universidad y el aprovechamiento del alumno se hacía una buena inserción laboral o si existían otras variables que participaban en el proceso.

Mediante la aplicación de conjuntos aproximados se descubrió que existían cuatro variables que determinaban la adecuada inserción laboral, que son citadas de acuerdo con su importancia: zona económica donde habitaba el estudiante, colegio de donde provenía, nota al ingresar y promedio final al salir de la carrera.

A partir de estos resultados, la universidad tendrá que hacer un estudio socioeconómico sobre grupos de alumnos que pertenecían a las clases económicas bajas para dar posibles soluciones, debido a que tres de las cuatro variables no dependían de la universidad.

Objetivo General

Establecer el perfil por segmento de los alumnos enfocado a los servicios de tecnologías de información de la UPeU, aplicando el algoritmo de K-medias.

Marco Teórico

Segmentación

La segmentación de mercado puede enfocarse en función a la metodología empleada y de los

objetivos buscados.

Segmentación de mercados a Priori: Centrándose en el primer enfoque, si se selecciona un criterio de división del mercado previamente a la recogida de información, clasificando posteriormente a los individuos en función de dichas variables

Segmentación de mercados Post-hoc: Clasifica a los individuos considerando su grado de similitud en función de un grupo de variables, posteriormente se estudia los perfiles de los grupos creados, según un segundo conjunto de variables, obteniendo de este modo una descripción de los segmentos. Además en el proceso de la división pueden quedar individuos sin clasificar. En este enfoque no se conoce previamente ni el número de grupos ni el tamaño de cada uno de ellos, ya que estos datos forman parte de los resultados de segmentación.

Finalmente la segmentación Híbrida: Es un planteamiento intermedio entre las dos anteriores. Divide el mercado en grupos de individuos utilizando datos relativos a variables, realizando en segundo lugar la descripción de dichos grupos mediante otras variables.

Acudiendo a la información básica de clustering, las técnicas de agrupamiento que se utilizará, dado que la clasificación es una de las abstracciones básicas en la inducción de conocimiento, de forma que el término agrupamiento se usa en numerosas comunidades de investigación para describir el proceso de clasificar en grupos.

Este estudio trabaja en base al tema de segmentación del mercado, que es el proceso de analizar el mercado con el fin de identificar grupos de clientes que tienen características comunes con respecto a la percepción de la calidad de los servicios de TI.

Bases para una correcta segmentación

La información acerca de las características del cliente debe ser fácilmente identificable.

El segmento de población seleccionada debe ser

fácilmente accesible.

El segmento debe ser medible. De lo contrario no sabremos qué recursos de ventas y de marketing dedicarles.

El segmento debe ser suficientemente diferente de otras partes del mercado con productos y precios, y así atraer solamente al segmento elegido.

El segmento seleccionado debe ser suficientemente grande para que sea significativo.

El segmento elegido debe ser duradero, no volátil, que ofrezca beneficios no sólo este año sino también los próximos.

Introducción a Clustering

En una primera aproximación, y acudiendo a la definición clásica, las técnicas de agrupamiento (clustering) se puede definir como: Una técnica no supervisada de elementos (observaciones, datos o vectores de características) en grupos (clústeres).

Dado que la clasificación es una de las abstracciones básicas en la inducción de conocimiento, este problema ha sido tratado en muchos contextos y por investigadores de muchas disciplinas (Biología, Psicología, Análisis Económico, Sociología, etc.), de forma que el término agrupamiento se usa en numerosas comunidades de investigación para describir el proceso de clasificar en grupos, un conjunto de elementos sin tener una información previa acerca de su estructura. Estas comunidades tienen diferente terminología para describir los componentes del proceso y diversas metodologías para resolver los problemas que el agrupamiento presenta.

Los problemas y metodologías para el agrupamiento pueden clasificarse según distintos criterios. En la figura 1, tomada de [Jain y otros, 1999] y [Lance y Williams, 1967], se puede ver un árbol con los distintos tipos de técnicas de agrupamiento. Se comentará brevemente esta

clasificación, y otras que se relacionan y superponen a ellas, ya que constituyen el hilo conductor de este capítulo:

El método de las k-medias

El algoritmo K medias se trata de un método de agrupamiento por vecindad en el que se parte de un número determinado de prototipos y de un conjunto de ejemplos a agrupar, sin etiquetar. Es el método más popular de los métodos denominados “por partición”. La idea del K medias es situar a los prototipos o centros en el espacio, de forma que los datos pertenecientes al mismo prototipo tengan características similares.

Considerando k-medias, se logra mediante iteraciones identificar los grupos asociados a un cluster, bajo los cuales se agrupan como se muestra en la Figura 1:



Figura 1 - Árbol de tipos de métodos de clasificación

Las regiones se definen minimizando la suma de las distancias cuadráticas entre cada vector de entrada y el centro de su correspondiente clase, representado por el prototipo correspondiente. (Figura 2).

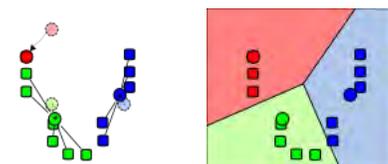


Figura 2 – Agrupamiento Jerárquico

Un enfoque diferente y más general para obtener agrupamientos jerárquicos es debido a Johnson [Johnson, 1967] y se basa en sucesivas transformaciones de la matriz de proximidad,

reduciendo la dimensión de la misma siempre que se forme un nuevo grupo. La idea consiste en trabajar con una matriz de distancia entre grupos, y que ésta se vaya calculando iterativamente a partir de la matriz de la etapa anterior.

Perfil de Usuarios y/o Alumnos

Constituye el elemento básico para diseñar servicios de información. Para precisar los perfiles de usuarios y/o clientes se debe seguir un proceso denominado Estudios de Usuarios, donde se enfatiza la importancia de los perfiles como fundamento para planear cualquier servicio, unidad o sistema de información (RefE3).

La mayoría de perfiles de usuarios están basados en las supuestas necesidades de información que los administradores creen detectar en los usuarios, sin probar o desaprobar dichos supuestos.

Los objetivos específicos de los estudios de usuarios dependerán de las características a analizar, entre las cuales podrían ser:

- Detectar las necesidades de información del usuario.
- Evaluar a satisfacción de usuarios.
- Detectar la actitud hacia la unidad o servicio de información.

Características del Usuario

Debe ser sistematizada y consistente, a través de la precisión de su perfil. El término perfil se deriva de la psicología, dentro de esta disciplina es entendido como el conjunto de medidas diferentes de una empresa o grupo, cada una de las cuales se expresa en la misma unidad de medición.

Ciertas características de un individuo son medidas mediante pruebas que arrojan puntuaciones diferentes, estas puntuaciones constituyen su perfil.

Datos que conforman un perfil de usuario

Disciplinas de interés (necesidades de información).

Niveles de educación.

Función o actividad principal: alumno, cliente, usuario, etc.

Recursos de información utilizados o requeridos.

Medios para localizar la información: citas, uso de los laboratorios, universidad, documentación personal, etc.

Comportamiento en la búsqueda de información.

Manejo del lenguaje.

La mayoría de perfiles de usuarios están basados en las supuestas necesidades de información que los administradores creen detectar en los usuarios, sin aprobar o desaprobar dichos supuestos.

Para comprobar o no las inferencias de perfiles de usuarios se deben llevar a cabo Estudios de Usuarios, estos estudios nos arrojarán el perfil de los usuarios o comunidades seleccionadas.

Técnicas para realizar Estudios de Usuarios

Las técnicas para recabar datos corresponden a las técnicas de investigación social y bibliotecológica:

Bibliografía: Comprende el análisis e interpretación de los registros que se consideren relevantes sobre el tema.

Encuesta: Es la técnica de campo, cuyo objetivo es obtener información que permita describir diversas facetas o situaciones de la comunidad bibliotecaria a estudiar. Existen instrumentos para recabar datos en una investigación por encuesta, que pueden ser:

Cuestionarios

Entrevistas

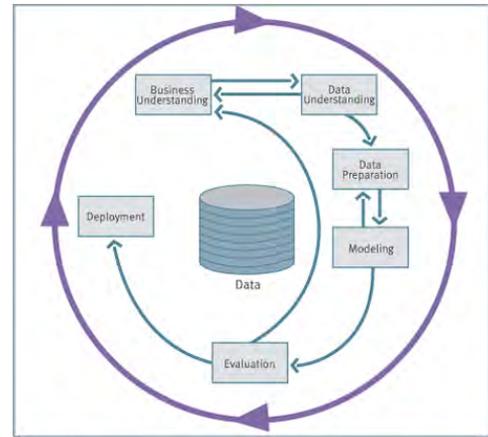
Observación directa

El usuario debe ser el agente principal para planear y/o desarrollar servicios de información. La infraestructura del conocimiento que un usuario tenga sobre una determinada disciplina, así como su saber empírico, determinarán más necesidades y uso que haga de la información.

Las necesidades de información, así como otras características del usuario, deben ser conocidas por el bibliotecario a través de la delimitación de perfiles. (RefE3)

Metodología de Estudio

La metodología usada fue la CRISP-DM de minería de datos (RefE1)



**Figura 3- Fases de minería de datos CRISP-DM
Procesos del Modelo**

Las fases de esta metodología son las siguientes:

Entendiendo el Negocio

Entendiendo los Datos

Preparación de los Datos

Modelamiento

Evaluación

Despliegue

Esta metodología fue acoplada a nuestro proyecto con todas sus fases respectivas en el siguiente modelo de solución.

La metodología CRISP-DM según Chapman, P (1999), consiste en un conjunto de tareas descritas en cuatro niveles de abstracción: fase, tarea genérica, tarea especializada, e instancia de proceso, organizados de forma jerárquica en tareas que van desde el nivel más general hasta los casos más específicos.

Fase: Se le denomina fase al asunto o paso dentro del proceso. CRISP-DM consta de 6 fases: comprensión del negocio, comprensión de los datos, preparación de los datos, modelación, evaluación y explotación.

Tarea genérica: Cada fase está formada por tareas genéricas, o sea, la tarea genérica es la descripción de las actividades que se realizan dentro de cada fase. Por ejemplo, la tarea Limpiar los datos es una tarea genérica.

Tarea especializada: La tarea especializada describe cómo se pueden llevar a cabo las tareas genéricas en situaciones específicas. Por ejemplo, la tarea Limpiar los datos tiene tareas especializadas, como limpiar valores numéricos, y limpiar valores categóricos.

Instancias de proceso: Las instancias de proceso son las acciones y resultados de las actividades realizadas dentro de cada fase del proyecto.

Las fases del proyecto de Minería de acuerdo a lo establecido por la metodología CRISP-DM interactúan entre ellas de forma iterativa durante el desarrollo del proyecto. La secuencia de las fases no siempre es ordenada, o en ocasiones si se determina al realizar la evaluación que los objetivos del negocio no se cumplieron se debe regresar y buscar las causas del problema para redefinirlo.

Herramientas

Herramientas de Software Estadístico
SPSS. Si bien SPSS es uno de los paquetes comerciales más conocido dentro del ámbito estadístico, cabe también destacar las herramientas que oferta para realizar agrupamientos. Éstas permiten realizar agrupamientos tanto desde el punto de vista más clásico, como algunas versiones más avanzadas. Estas herramientas son completas desde el punto de vista, funciones de distancia que se contemplan o de tipos de atributos que se consideran.

Herramientas de Software de Minería de Datos
CLEMENTINE. es una herramienta para la Minería de Datos desarrollada por SPSS, tiene un enfoque muy distinto al paquete estadístico. Posee un lenguaje de tipo gráfico muy intuitivo y permite ir almacenando resultados intermedios que se usan en procesos posteriores. En este sentido resulta una herramienta de análisis muy

cómoda. No obstante, desde el punto de vista de la Minería de Datos, la oferta de posibles técnicas es bastante pobre. Concretamente desde el punto de vista del agrupamiento oferta el método de las k-medias, el método de Kohonen, basado en redes neuronales, y el método de las dos fases, es decir de BIRCH que ya se oferta también en SPSS, no hay agrupamiento jerárquico, ni métodos de medoides (RefE2).

Es una herramienta integrada de minería de datos, inicialmente de Integral Solutions Limited (ISL) y ahora de SPSS (www.spss.com).

La versión 5.2.1 en inglés o la 6.0.2 en castellano, que es con las que vamos a trabajar indistintamente, incluye las siguientes características:

Interfaz visual basado en procesos/flujo de datos (streams).

Distintas herramientas de minería de datos: correlación, reglas de asociación (GRI, a priori).

patrones secuenciales (regresión), segmentación (Kohonen, Two-step y k-means), clasificación

Manipulación de datos (pick & mix, muestreo, combinación y separación).

Combinación de modelos.

Visualización anterior (datos).

Exportación de modelos a distintos lenguajes (C, SPSS, SAS).

Exportación de datos integrada a otros programas (XLS).

Generación de informes.

Clementine cuenta con características flexibles y amigables que le permitirán ahorrar tiempo y proporcionarle a su empresa resultados analíticos

de gran impacto para el negocio (Hernández J, 2002).

Resultados y discusión

Por lo tanto se procedió a analizar las características de cada uno de estos conglomerados y como resultado de ello se dieron los siguientes nombres. La clasificación fue por conglomerados como se muestra en la tabla 4.

Tabla 4 - Clasificación por conglomerados

Encuestado	\$KM-K-Medias	\$KMD-K-Medias
1	conglomerado-1	1.7834311389299642
2	conglomerado-3	1.4493402170090155
3	conglomerado-3	1.8975216140612314
4	conglomerado-2	0.7184734314032567

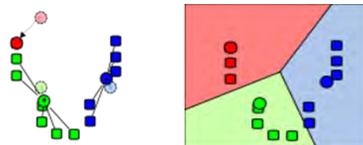


Figura 4- Agrupamiento Jerárquico

Y de estos se encontró 3 grupos de mayor relevancia.

Grupo Terrorista: Grupo de alumnos que están descontentos por todo, como resultado de ello, pueden sembrar el descontento con sus compañeros e incluso al salir de la institución pueden hablar mal de los servicios que ofrece la UPeU.

Las características del grupo terrorista es que para ellos todos los servicios de TI no están al nivel de lo que ellos esperan, no aprueban ni siquiera un servicio en específico de los 13 servicios específicos que se ha evaluado y por este mismo hecho es de esperarse que ellos siembren este descontento tanto a sus demás compañeros como a terceros a la institución.

Grupo Realista: Este grupo trata de equilibrar de manera más concienzuda la evaluación de los servicios de TI, es muy probable que en muchos aspectos tengan toda la razón, pues no llevan a los extremos de sentirse tan satisfechos por todos los servicios, ni tampoco el otro extremo de

sentirse descontentos. Si bien es cierto hay servicios que ellos creen que necesitan ser mejorados, también sienten satisfacción por los demás servicios de TI.

El grupo realista es más equilibrado, de los 13 servicios que se evaluaron, aprueban a 5 y 8 son servicios críticos que para ellos necesitan ser mejorados, de estos 8 servicios, 4 corresponden a los servicios de internet que abarca los siguientes servicios específicos: Velocidad del internet, Restricciones del internet, Disponibilidad de los laboratorios de internet, Horarios de los servicios de internet. Así mismo se tienen como servicios críticos a los siguientes: Atención en los laboratorios de cómputo, Contenido de la web UPeU, navegación en el portal académico y web UPeU, Seguridad del portal académico, entre otros puntos de mayor atención para con los servicios brindados por la universidad.

Grupo Apóstol: Lo opuesto del grupo terrorista, están casi completamente satisfechos con todos los servicios de TI. Este grupo siempre hablará bien de los servicios de la institución vayan donde vayan y con ello promoverán para dar una buena imagen de su universidad, especialmente al salir de ella.

El servicio de Internet necesita mejorar en cuanto a su velocidad, es decir, la gran mayoría está descontento con este servicio, aun el grupo apóstol que se caracteriza por estar contento con el resto de servicios de TI, el grupo apóstol está insatisfecho en un 75% lo cual dada las características del grupo es importante considerarlo, pues no podría ser un invento del grupo terrorista por desprestigiar a la institución.

Puntos críticos de los servicios de TI en la UPeU

Considerando de manera global, en los tres grupos coinciden en que el servicio de internet necesita, es decir, la gran mayoría está descontento con este servicio, aun el grupo apóstol que se caracteriza por estar contento con

el resto de servicios de TI, el grupo apóstol está insatisfecho en un 75% como se menciona anteriormente.

De los tres conglomerados, se tiene que el conglomerado 1 (grupo terrorista) equivale al 37.4%. De ellos está mayormente insatisfecho con el servicio de internet (97.67%), le sigue las conexiones e instalaciones de los laboratorios de cómputo y la seguridad de los mismos (95.35%), así mismo la atención por parte del personal de los laboratorios de cómputo es un punto crítico con alto porcentaje (93.02%), continúa la seguridad del portal académico, el cual según ellos en un 90.7% no es fiable en la seguridad. Otros servicios de TI que tienen un nivel alto de insatisfacción es con la disponibilidad de los laboratorios de internet, con los horarios de los servicios de internet y con las restricciones del internet que incluyen páginas que están restringidas (88.37%). En este último punto es rescatable que la UPeU restringe páginas que no son académicas, tales como páginas de redes sociales y de chat, se podría interpretar que los alumnos desean éstas páginas que estén disponibles, pero al no estar se siente un alto nivel de insatisfacción; que por el bien de ellos, para no desconcentrarse de los temas de clases se les ha restringido.

Del conglomerado 2 (grupo apóstol) equivale al 35.7% que se identifica porque están satisfechos y aprueba casi todos los servicios de TI que brinda la UPeU, a excepción de velocidad del internet que es un punto en común de los tres conglomerados. De ellos el servicio que más aprueban es el diseño de la web UPeU (97.56%), le siguen conexiones e instalaciones de los laboratorios de cómputo y navegación la web UPeU y portal académico (92.68%), seguidamente del ambiente de los laboratorios de cómputo y seguridad de las mismas (87.8%), el servicio que aprueban con menor porcentaje, pero que pasa el 50%, es disponibilidad de los laboratorios de cómputo (56.1%) lo cual significa que lo desaprueban en un 43.9%.

El conglomerado 3 (grupo realista) que equivale al 27% del total, de ellos el servicio que más desaprueban es la velocidad del internet (90.32%),

seguido de horarios de los servicios de internet que lo desaprueban en un 87.1%, también desaprueban el contenido de la web UPeU en un 74.19%. Los servicios que aprueban en mayor porcentaje es el ambiente de los laboratorios de cómputo (93.55%), le sigue conexiones e instalaciones de los laboratorios de cómputo (87.1%), después le sigue diseño de la página web (74.19%). El resto de servicios para este grupo están casi equilibradas, es decir van casi 50/50 tanto para aprobarlos y considerarlo como puntos críticos que necesitan ser satisfechos.

Conclusiones

Se obtuvieron los siguientes perfiles de alumnos:

Considerando de manera global, en los tres grupos coinciden en que el servicio del internet necesita mejorar en cuanto a su velocidad, es decir, la gran mayoría está descontento con este servicio, aun el grupo apóstol que se caracteriza por estar contento con el resto de servicios de TI, el grupo apóstol está insatisfecho en un 75% lo cual dada las características del grupo es importante considerarlo.

Las características del grupo terrorista es que para ellos todos los servicios de TI no están al nivel de lo que ellos esperan, no aprueban ni siquiera un servicio en específico de los 13 servicios específicos que se ha evaluado y por este mismo hecho es de esperarse que ellos siembren este descontento tanto a sus demás compañeros como a terceros a la institución.

El grupo realista es más equilibrado, de los 13 servicios que se evaluaron, aprueban a 5 y 8 son servicios críticos que para ellos necesitan ser mejorados, de estos 8 servicios, 4 corresponden a los servicios de internet que abarca los siguientes servicios específicos: Velocidad del internet,

siguientes servicios específicos: Velocidad del internet, Restricciones del internet, disponibilidad de los laboratorios de internet, Horarios de los servicios de internet. Así mismo se tienen como servicios críticos a los siguientes: Atención en los laboratorios de cómputo, Contenido de la web UPeU, Navegación en el portal académico y web UPeU, Seguridad del portal académico.

|Recomendaciones

Considerando de manera global, en los tres grupos coinciden en que el servicio del internet necesita mejorar en cuanto a su velocidad, dicho anteriormente. Si bien es cierto que algunos laboratorios no tengan buen ancho de banda, es probable que los alumnos confundan la velocidad de internet con la capacidad de las maquinas porque, sería necesario identificar bien cuál es la causa precisa de esta necesidad con algún tipo de test de las maquinas y de la red.

Agradecimientos

Agradecemos a todas las personas involucradas a que este proyecto pueda ser posible, amigos y profesores. Damos un agradecimiento especial al Dr. Edgardo Palza por el apoyo en la fase de desarrollo e implementación. Y a los siguientes alumnos del 5to año de ingeniería de sistemas de la Upeu por su apoyo en la revisión y mejora de este artículo:

- Jenrry Chuquista T.
- Jorge Llampén C.
- Moshe Cotacallapa

Referencias.

Brachman R, Khabaza T, Kloesgen W, Piatetsky-Shapiro G, Simoudis E. (1996). "Minería de datos en negocios". Comunicaciones del ACM. Vol. 39, pág.42-48.

Chapman P, Clinton J, Kerber R, Khabaza T, Reinartz T, Shearer C, Wirth R. (1999). CRISP-DM 1.0 Guía de Minería de datos paso a paso.

Escobar V, 2007, Minería web de uso y perfiles de usuario: aplicaciones con lógica difusa. Granada. TESIS DOCTORAL.

Rodas, J. (2001). "Un ejercicio de análisis utilizando rough sets en un dominio de educación superior mediante el proceso KDD". Documento interno. Barcelona: Departamento de Lenguajes y Sistemas Informáticos, Universidad Politécnica de Cataluña.

Referencias Electrónicas

(RefE1)Cross Industry Standard__Process for Datamining(CRISP-DM). Modelo del proceso. [Consultado en 20 de agosto del 2010]. Disponibilidad libre en: <<http://www.crisp-dm.org/Process/index.htm>>

(RefE2) Clementine especificaciones. [Consultado en 15 de septiembre del 2010] . Disponibilidad libre en: <<http://www.spss.com/la/productos/clementine/clementine.htm>>

(RefE3)Perfil de alumnos y/o usuarios en la Empresa en una entidad. Disponibilidad libre en:<<http://www.actiweb.es/ipgcrp/pagina4.html>. Perfil de alumnos y/o usuarios en la Empresa en una entidad>