

# Model Clustering Neural Network Monitoring Applications No K-Means Algorithm for the Segmentation of the Peruvian University Students Union

Rodolfo Pacco, Palomino; Del Rosario Montero, Wilson

Tomas Anquise, Ibañiz

Dr. Palza Vargas, Edgardo; Dr. Mamani Apaza, Guillermo.

## Resumen

El presente trabajo de investigación tiene como objetivo identificar el perfil del alumno moroso utilizando algoritmo de clústeres basado en la red neuronal de k-media para la segmentación de los alumnos de la UPeU, de manera que exista la mayor homogeneidad posible dentro de los grupos, con respecto a ciertos atributos que pueden ser cuantitativas o cualitativas. La metodología utilizada fue CRISP-DM, que es un estándar para proyectos de minería de datos, consta de las siguientes 6 fases, la comprensión del negocio, la comprensión de los datos, la preparación de datos, el modelado, la evaluación e la implementación. Con la aplicación del algoritmo de inteligencia artificial se ha determinado, que el perfil de un alumno moroso tiene las siguientes características, el alumno que no trabaja, el ingreso bruto del apoderado varía entre S/850.00 y S/1,569.00, tiene hermanos estudiando en la misma universidad y está cursando el tercer año de su carrera.

**Palabras claves:** Clusters, Crisp-DM, Morosidad, Sigmoidal, Algoritmos, K-medias, UPeU.

## Classification Trees Model for Identifying the Student's Profile according to Credit Risk of Universidad Peruana Unión

### Abstract

The present research aims to identify the delinquent Student profile clustering algorithm using neural network based on k-means for segmenting the UPeU students, so that there is greater homogeneity within the groups may, with respect to certain attributes that can be quantitative or qualitative.

The methodology used was CRISP-DM, which is a standard for data mining projects, has the following 6 phases, the understanding of Negotiating, data understanding, data preparation, modeling, evaluation and implementation.

With the application of artificial intelligence algorithm has determined that the profile of a delinquent student has the following characteristics, the student who does not work, the gross income of the agent varies between S/850.00 and S / 1,569.00, has siblings studying in the same university and is in his third year of his career.

**Keywords:** Clusters, Crisp-DM, Delinquency, sigmoidal, Algorithms, K-means, UPeU.

## 1. Introducción

Según **Raposo, J. M. (2009)**, el impacto del riesgo de crédito en las entidades se refleja, además de un aumento de la morosidad, en mayores necesidades de dotaciones a insolvencias. Por otro lado en el primer tratado internacional de **Basilea I en 1988 se calificaron estos riesgos crediticios en 5 niveles:**

– Categoría A o “Riesgo normal”.

- Categoría B o “Riesgo aceptable”.
- Categoría C o “Riesgo apreciable”.
- Categoría D o “Riesgo significativo”.
- Categoría E o “Riesgo de incobrabilidad”

La gerencia de finanzas no tienen el conocimiento de segmentación de alumnos por grupos de morosidad, al momento de matricularse, el alumno asume pagar sus pensiones de enseñanza donde es dividida en 5 partes o cuotas.

La condición o situación de cada alumno es imposible conocer por parte del personal de finanzas ya que el sistema actual de finanzas no reporta.

Hoy en día la mayoría de los estudiantes se auto sustentan o son de pocos recursos económicos, no cuentan con apoyo de sus padres, en algunos casos los alumnos reciben ayuda puntual de sus padres sin embargo el alumno simplemente no paga su pensión de enseñanza por motivos que se conocer los factores que causan la morosidad de los alumnos.

El objetivo del algoritmo k-medias clustering es reducir la cantidad de datos mediante la caracterización o agrupamiento de datos según las características similares. Usar algoritmos k-medias clustering es proveer herramientas automáticas que ayuden a la construcción de taxonomías,[4].

El algoritmo k-medias que está incluido dentro de la herramienta de Microsoft, es para la segmentación suministrado por SQL Server 2008 Analysis Services (SSAS). El algoritmo utiliza técnicas iterativas para agrupar los casos de un conjunto de datos dentro de clústeres que contienen características similares. Estas agrupaciones son útiles para la exploración de datos, la identificación de anomalías en los datos y la creación de predicciones.

## 2. Morosidad.

**2.1. Moroso:** Se considera a los alumnos que se atrasan o tienen deficiencia de pago de sus pensiones de enseñanza, según las cuotas pactadas durante la matrícula.

**2.2. No moroso:** Se considera a los alumnos que pagan sus pensiones de enseñanzas de forma puntual, ya sea al contado o divididas en 5 cuotas pactadas durante la matrícula.

## 3. Materiales y método.

**3.1. SQL server 2008:** El algoritmo de k-medias clustering es el la herramienta de SQL Server 2008 incluye el algoritmo de K-medias clustering

para segmentar en grupos, el último lanzamiento de Microsoft SQL Server, ofrece una plataforma de datos completa, más segura, confiable, administrable y escalable para aplicaciones críticas. Permite que los desarrolladores creen aplicaciones nuevas, capaces de almacenar y consumir cualquier tipo de datos en cualquier dispositivo, y que todos los usuarios tomen decisiones informadas en base a conocimientos relevantes.

### SQL Server ofrece:

Los niveles más altos de seguridad, confiabilidad y escalabilidad para aplicaciones de negocios críticas.

Una sólida plataforma, proveedora de inteligencia justo donde sus usuarios la necesitan.

**3.2. Metodología Crisp – DM:** La metodología Crisp – DM, es un estándar para la realización de proyectos de minería de datos, en donde el ciclo de un proceso de minería de datos está estructurado en 6 fases, como se puede observar en la figura Nro. 1, algunas de estas fases son bidireccionales, lo que significa que algunas fases permitirán revisar parcial o totalmente las fases anteriores, [1].



Figura 1 - Modelo de la metodología de CRISP-DM.

**Comprensión del negocio.** Se ha realizado la entrevista con el personal de finanzas alumnos para diagnosticar la situación actual en la que se encuentra finanzas alumnos de la Universidad

Peruana Unión

**Comprensión de los datos.** De los datos obtenidos se seleccionaron los más relevantes, tales como responsable financiero, tipo de beca, número de hermanos y situación laboral del alumno, para desarrollar el proyecto.

**Preparación de los datos.** Se elaboró cuestionario para la obtención de datos, y se procedió a encuestar a la población de un total de 800 alumnos.

**Modelado.** Utilizando el algoritmo K-medias que está incluido dentro de la herramienta SQL Server 2008 Analysis Services, se agrupó los datos obtenidos según las características.

**Evaluación.** Se evaluó dos grupos moroso y no moroso, haciendo la extracción y análisis de datos utilizando el algoritmo K-medias dentro de la herramienta de Microsoft.

#### 4. Resultado y discusión.

El modelo de Redes Neuronales Artificiales basados en algoritmo K-medias, en la figura Nro. 2, se presentan los grupos que arrojó el modelo en función a la morosidad, y son cuatro clusteres: No moroso, poco morosos, Moroso y muy morosos.

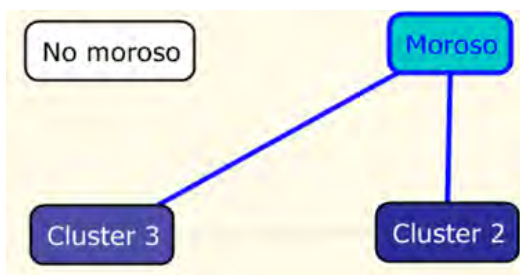


Figura 2 - Grupos en función a la morosidad

**Análisis de atributo financiamiento:** El atributo más importante en el clúster "Moroso" como se puede observar en la figura Nro. 3 es el financiamiento, eso quiere decir que el 81.1% del total de los alumnos morosos, tienen como responsable de pago a sus padres, y en el clúster "No moroso", el responsable de pago de estos

alumnos es el padre y en un 70% además los mismos estudiantes se apoyan trabajando en un 30%.

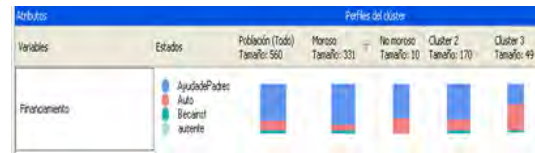


Figura N. 3 - Ficha Perfiles del clúster para el atributo Financiamiento.

Se muestra en la figura Nro. 4, que el 70.00% de los 560 encuestados recibe ayuda de sus padres, el 14.62% de los 116 alumnos encuestados se auto sostienen, 8.46% de los 67 encuestados recibe ayuda de becas y el 6.92% de los 56 encuestados recibe Beca ley.

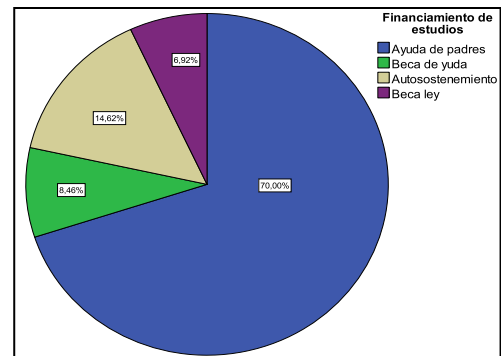


Figura Nro. 4 – Financiamiento de estudios.

**Análisis del atributo Situación laboral:** Como podemos observar en la figura Nro. 5, en el clúster "Moroso", el 67.5% no trabaja, y en el otro clúster "No moroso" el 100% trabaja.



Figura Nro. 5 - Ficha Perfiles del clúster para el atributo Situación Laboral.

**Análisis del atributo Hermanos en la Universidad:** En la siguiente figura Nro. 6, el clúster "Moroso" tiene en promedio 1 hermano estudiando en la universidad, y en el clúster "No moroso" no tiene hermanos estudiando en la universidad.

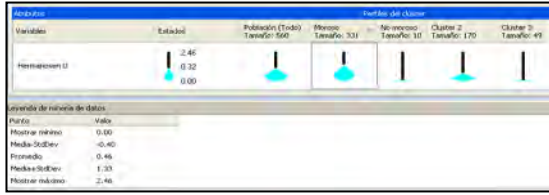


Figura Nro. 6 - Ficha Perfiles del clúster para el atributo Hermanos U.

Se muestra en la Figura Nro. 7, que 63.31% de los 506 encuestado no tienen ningún hermano estudiando en la UPeU, el 17.69% de los 142 encuestados tienen un hermano estudiando en la universidad, el 14.62% de los 117 encuestados tiene 2 hermanos estudiando y los 5.38% de los 43 encuestados tienen tres hermanos estudiando en la universidad. Esto quiere decir que los estudiantes que tienen más de un hermano estudiando en la universidad, es muy probable que tenga deficiencias de pago de sus cuotas de mensualidad.

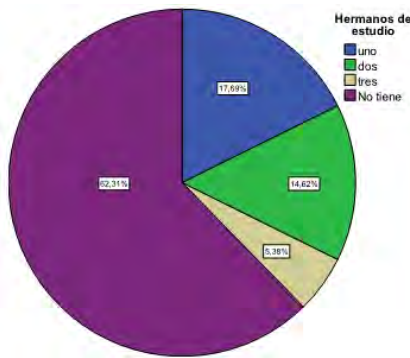


Figura Nro. 7 - Resultados obtenidos hermanos de estudio.

**Clúster moroso:** Según las fichas de características de clúster en la figura Nro. 8, la probabilidad de que el alumno sea moroso si es que estudia en otra institución es 84.211%, el alumno sea moroso y el responsable del pago sea su padre es de 81.115%, una variable que favorece a la morosidad es si el alumno no trabaja, por lo tanto la probabilidad de que el alumno no trabaje influya en la morosidad es de 67.492%, y la probabilidad de que sea moroso dado que el cargo laboral de su apoderado es empleado es de 69.659%.

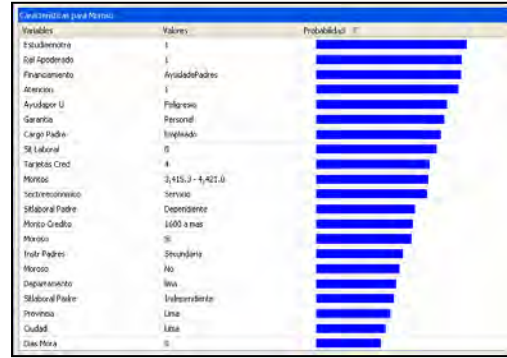


Figura N. 8 - Ficha Características del clúster.

**Clúster no moroso:** Según las ficha distinción del clúster en la figura Nro. 9, la probabilidad de que el alumno no sea moroso dado que trabaja es de 100.00%, el alumno no sea moroso dado que su apoderado tiene instrucción secundaria es de 100.00%, el alumno no sea moroso dado que su apoderado tiene un ingreso bruto entre S/ 1010.5 y 1432 es de 100.00% y la probabilidad de que el alumno no sea moroso dado que tiene apoyo de sus padres es de 70.00%.

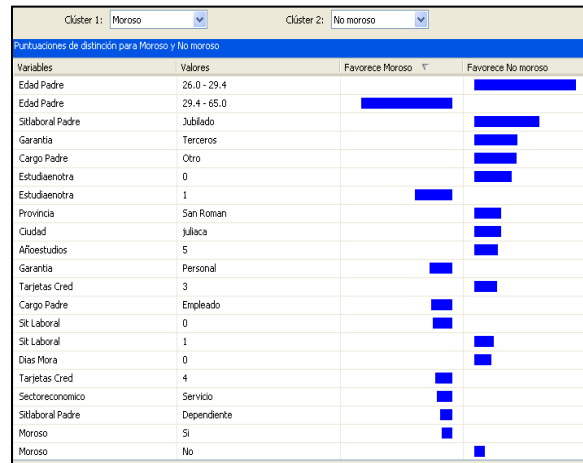


Figura N. 9 - Ficha distinción del clúster.

## 5. Conclusiones.

Para poder conocer el perfil del alumno moroso fueron identificados las siguientes características: el alumno que trabaja y recibe apoyo de sus padres es el alumno que tiene mayor probabilidad de que no sea moroso, el ingreso del apoderado es la característica más resaltante en ambos clústeres, el ingreso promedio del apoderado que no se retrasa en las cuotas y su hijo es no moroso es de S/ 1,285.00, la mayoría de los

alumnos morosos pertenecen al 3ro año de la carrera.

En resumen, la investigación realizada aporta evidencia empírica a favor de la especificación u obtención de modelos de segmentación para finanzas alumnos de UPeU, con el fin de evaluar el riesgo de morosidad.

### 6. Recomendaciones.

A la gerencia de finanzas alumnos se recomienda la segmentación de alumnos según las características cas de morosidad para tener reportes específicos, que le ayuden a tomar decisiones.

Continuar con esta propuesta de investigación que podría dar lugar a constituirse en una línea de investigación. Asimismo, es oportuno conformar en nuestra Universidad grupos de alumnos que desarrollen la implementación de un modelo de clusterización K-medias con el objetivo de agrupar según las características.

### Referencias

- [1] Crisp - DM. 2007. Cross Industry Standard Process for Data Mining, Fundación Europera. <<http://www.crisp-dm.org/index.htm>>
- [2] Lara G. 2008. La Técnica del árbol para la toma de decisiones. México: Univalle. 350p.
- [3] Martínez González, D. (2004-2005). Redes Neuronales Artificiales y Mapas Auto Organizados. Curso (Sistemas Expertos e Inteligencia Artificial) 2004-2005. Ciudad Universitaria de Burgos, Burgos.
- [4] Microsoft. (s.f.). microsoft /TechNet. Recuperado el 15 de mayo de 2010, de Libros en pantalla de SQL Server 2008 R2, Consultar un modelo de clusters (Analysis Services - Minería de datos). Disponible en: <<http://technet.microsoft.com/es-es/library/cc280440>>
- [5] Raposo, J. M. 2009. *Estimación de un modelo de calificación de pequeñas y medianas*

*empresas para evaluar el riesgo de crédito.* Tesis Doctoral, Universidad Complutense de Madrid. 382 pp. Disponible en: <<http://eprints.ucm.es/9631/1/T31333.pdf>>

### Agradecimientos

Agradecemos sinceramente al grupo de estudiantes conformado por:

- Giovedy Marmolejo Cruz.
- Obrian Quinto Velarde.
- José Luis Matías Villalobos.
- Teddy Huillca Jaimes.

Quienes revisaron el presente artículo, y colaboraron con el mejoramiento del mismo.