

Evaluación de técnicas de data mining para la obtención de perfiles de inscriptos a la UNNOBA

La UNNOBA recolecta información a través de sus sistemas de gestión provenientes de la inscripción a las propuestas académicas de grado y posterior desarrollo del curso de ingreso. Esta información puede ser transformada en conocimiento ya sea para la toma de decisiones que conciernen a la población estudiantil, como para el diseño de planes de acción.

El proceso de descubrimiento de conocimiento (KDD) incluye: comprensión del problema, análisis exploratorio de los datos, reducción de la dimensión del problema, selección del algoritmo que produzca los patrones más expresivos, validación de la solución obtenida.

En etapas anteriores del actual proyecto se obtuvo una vista minable de los datos extraídos del Sistema de Preinscripción a la Universidad. Luego se redujo la dimensión del problema y se ejecutaron distintos algoritmos de clustering. Se llegó a la conclusión que haciendo uso de las variables y el coeficiente de disimilitud de Gower no es posible llegar a los resultados esperados.

El objetivo de la presente beca PROMINF es: reducir la dimensión del problema mediante distintos métodos, evaluar algoritmos de clustering haciendo uso de distintas medidas de distancia, ensamblar con algoritmos que describan y expliquen mejor el conocimiento obtenido y validar la solución.

La metodología a aplicar será la siguiente: convertir las variables a dummies; evaluar: técnicas de selección de características, técnicas para obtener modelos descriptivos, el ensamble de algoritmos, técnicas de visualización de los resultados. Documentar las herramientas producidas y los resultados observados de la experiencia, proponer como futuros pasos nuevos objetivos de investigación.

Hasta el momento se evaluaron algoritmos de clustering usando distintas medidas de distancia. Los resultados obtenidos no fueron los esperados, se decide aplicar algoritmos de aprendizaje supervisado.

Palabras claves: Deserción, Minería de datos educacional, Algoritmos de Clustering.