

UNIVERSIDAD NACIONAL DE SAN ANTONIO ABAD DEL CUSCO
FACULTAD DE INGENIERÍA ELÉCTRICA, ELECTRÓNICA, INFORMÁTICA Y
MECÁNICA
ESCUELA PROFESIONAL DE INGENIERÍA INFORMÁTICA Y DE SISTEMAS



TESIS

**“ OPTIMIZACIÓN DE MÉTODOS RADIALES DE EXPLORACIÓN DE
DATOS MULTIDIMENSIONALES A TRAVÉS DE ALGORITMOS DE
APRENDIZAJE MÁQUINA ”**

Para optar al título profesional de:
INGENIERO INFORMÁTICO Y DE SISTEMAS

Presentado por:
BR. GERMAIN GARCÍA ZANABRIA

Asesor:
DR. LAURO ENCISO RODAS

Cusco - Perú
2017

Resumen

El descubrimiento de patrones en conjuntos de datos despierta gran interés en la comunidad de investigadores. Específicamente, en el manejo de datos multidimensionales o datos de alta dimensión se han realizado grandes avances. Sin embargo, todavía existen ciertas limitaciones como: costo computacional, falta de percepción visual y la interacción con el usuario que impiden que la tarea de extracción de patrones y agrupamientos sea simple y eficiente. El grupo de técnicas que mitigan estos problemas son los denominados *Técnicas de Visualización Radial* (*Star Coordinates*, *Radviz*, *Barycentric Coordinates*), los cuales son capaces de revelar patrones y grupos de datos multidimensionales mientras muestran el impacto de los atributos en la formación de la representación de los datos. A pesar de su utilidad, las técnicas de exploración radial tienen ciertos inconvenientes que impiden su uso en varios escenarios. Por ejemplo, cuando el número de dimensiones de los datos es realmente alto, las visualizaciones resultantes se vuelven desordenadas y sobrecargadas, lo que dificulta el análisis de la importancia de los atributos en la formación de grupos y/o patrones. En este trabajo de investigación se optimiza las técnicas de visualización radial aprovechando sus ventajas (interacción y bajo costo computacional) y fortaleciendo sus deficiencias (manejo de datos con un número alto de dimensiones) en el análisis de datos multidimensionales. La optimización de estas técnicas se cimienta en el agrupamiento (similaridad de atributos) y reordenamiento de atributos (optimización de ubicación) con el fin de mitigar la distorsión visual. El agrupamiento y reordenamiento se puede realizar de forma automática, así como de forma interactiva, lo que permite que el usuario pueda analizar aún más el impacto de los atributos en la visualización radial. La eficacia de los enfoques presentados se muestran a través de una serie de experimentos y estudios de caso, los cuales evidencian utilidad de la propuesta.

Palabras clave: Visualización de Información, Datos Multidimensionales, Star Coordinates, Radviz, Barycentric Coordinates, agrupamiento y reordenamiento de atributos.