Resultados

Este capítulo muestra los resultados obtenidos en cada fase de la metodología de esta investigación.

CRISP-DM .

Comprensión del problema: En esta fase se obtuvo como resultado un marco teórico referente a todas las técnicas utilizadas en esta investigación para tratar con documentos textuales no estructurados.

Preparación de los datos: En esta fase se obtuvo un corpus limpio compuesto por 10.000 documentos.

Modelado: En esta fase se obtuvo diferentes modelos para estructurar el conjunto de datos.

Conjunto de datos estructurados con los modelos –BOW de dimensiones (10.000, 20.000) y (10.000, 10.000), según el estadístico Hopkins de 0.61 para este conjunto de no existe tendencia al agrupamiento con un umbral establecido en 0.5

Selección del k optimo

para la selección del k óptimo generamos diferentes grupos iterando k desde 20 a 44 evaluando el coeficiente de Silhouette y el error cuadrático en los 2 conjuntos de datos generados con el modelo Doc2vec ya que tienen tendencia al agrupamiento.