

## Abstrak

Bahasa merupakan cara berkomunikasi antar manusia untuk saling memahami satu sama lain. Keragaman pola bahasa menyebabkan kurang terstrukturanya bahasa yang digunakan. Struktur bahasa dapat dikenali dari kelas kata yang dikandung. Proses pengenalan kelas kata dikenal dengan *POS tagging*. Dalam proses pengenalan kelas kata perlu dilakukan pengelompokan kata berdasarkan kedekatan antar katanya. Berdasarkan hal tersebut, metode SVD dan X-means *clustering* dapat digunakan untuk mengelompokkan kata-kata.

Metode SVD digunakan untuk memodelkan permasalahan kedekatan antar kata ke dalam matrik deskriptor. Matrik deskriptor mengandung nilai konteks kiri dan konteks kanan dari masing-masing kata di dalam corpus. Matrik deskriptor digunakan untuk proses *clustering* dengan metode X-means *clustering*. X-means *clustering* mengelompokkan kata-kata sesuai dengan *cluster* yang akan terus terbentuk hingga tercapai batas atas yang telah diatur. *Cluster* yang terbentuk diuji kualitasnya menggunakan *Silhouette Coefficient*.

Kombinasi SVD dan X-means *clustering* menghasilkan kualitas *cluster* dengan struktur baik yaitu menengah menggunakan inisialisasi *centroid sequence*. Parameter Top N1, RR1, Top N2, dan RR2 dalam metode mempengaruhi nilai SC yang didapatkan. Semakin besar selisih Top N dan RR maka nilai SC yang dihasilkan semakin baik dan kualitas *cluster* yang terbentuk semakin baik pula.

**Kata kunci:** *Part of speech tagging, SVD, X-means clustering, Silhouette Coefficient.*