

Abstrak

Klasifikasi merupakan proses mengelompokkan suatu data ke dalam kelompok data yang telah ditentukan berdasarkan tingkat kemiripannya. Klasifikasi ini pun dapat diterapkan dalam dokumen teks, dengan tujuan mempermudah penentuan seluruh dokumen dengan kategori tertentu. Terdapat berbagai cara untuk melakukan klasifikasi, salah satunya dengan menggunakan metode *K-Nearest Neighbor*. Metode *K-Nearest Neighbor* merupakan metode yang populer dalam klasifikasi, karena kemudahan dalam implementasinya.

Tetapi dibalik kemudahannya itu metode *K-Nearest Neighbor* memiliki kelemahan jika digunakan dalam dokumen yang memiliki distribusi yang tidak merata, karena saat nilai k yang digunakan semakin besar akan ada dominasi oleh kelas yang berukuran besar terhadap kelas yang berukuran kecil. Oleh karena itu digunakan metode *Improved K-Nearest Neighbor* untuk menanggulangi kelemahan tersebut.

Untuk mengevaluasi performansi dari *K-Nearest Neighbor* dan *Improved K-Nearest Neighbor* digunakan *precision*, *recall*, dan *F1-Measure*. Hasil yang didapat menunjukkan bahwa metode *Improved K-Nearest Neighbor* dapat menghilangkan efek dominasi dari kategori terbesar dalam berbagai jenis distribusi dokumen *training*.

Kata kunci: klasifikasi, *K-Nearest Neighbor*, *Improved K-Nearest Neighbor*