

1. Pendahuluan

1.1 Latar belakang masalah

Semakin hari, informasi yang berupa dokumen, teks atau artikel semakin banyak dihasilkan dan dibutuhkan oleh seluruh kalangan masyarakat. Sebagai contoh, artikel berita. Setiap hari jumlah artikel berita makin meningkat karena selalu ada berita terbaru. Jika peningkatan ini tidak dikelola dengan baik maka masyarakat tidak akan mendapatkan informasi yang diperlukan dengan cepat dan mudah.

Untuk mengatasi masalah tersebut maka muncullah suatu bidang ilmu yang disebut *text mining*. *Text mining* adalah proses pencarian informasi yang didapat dengan memproses data teks dalam jumlah yang besar yang ditulis dengan bahasa manusia. Aktivitas-aktivitas yang terdapat pada teks mining adalah *Information Retrieval*, *Text Categorization*, *POS Tagging*, *Clustering*, dll. Pembahasan dalam Tugas Akhir ini lebih fokus pada tahap *categorization*. Aktivitas pada *categorization* lebih terfokus pada penentuan kelas-kelas dari beberapa dokumen ke dalam kelas-kelas yang sudah didefinisikan sebelumnya.

Pada tahap *text categorization* atau klasifikasi teks terdapat banyak metode yang dapat digunakan. Namun demikian, tidak semua metode cocok untuk diterapkan pada berbagai masalah klasifikasi terutama klasifikasi teks dalam jumlah besar. Oleh karena itu, pada Tugas Akhir ini digunakan metode *Weight Adjusted k-Nearest Neighbor* (WAKNN). Metode ini merupakan perbaikan dari metode *k-Nearest Neighbor* yang merupakan salah satu algoritma learning yang sangat efektif untuk berbagai domain permasalahan. Namun, *k-Nearest Neighbor* (KNN) dianggap kurang efektif dalam pengukuran *similarity*/kemiripan sifat antara satu dokumen dengan dokumen yang lainnya karena memakai semua kosakata yang terdapat dalam dokumen tersebut baik penting maupun tidak.

Dalam hal ini, WAKNN akan menghitung dan mengevaluasi bobot kata dari setiap dokumen untuk menentukan kata-kata penting dari suatu kelas sehingga pada proses klasifikasi, setiap dokumen akan dibandingkan antara satu dengan yang lainnya sesuai dengan kata penting yang dimilikinya.

1.2 Perumusan masalah

Beberapa permasalahan yang akan dibahas pada Tugas Akhir ini adalah:

1. Bagaimana mengubah suatu dokumen berita menjadi data yang siap untuk digunakan dalam proses klasifikasi melalui tahap *preprocessing* yang meliputi: proses penghilangan *stopword*, dan *feature selection*.
2. Bagaimana proses klasifikasi data yang berupa teks berita dengan menggunakan WAKNN sehingga diperoleh suatu model yang akurat dalam proses klasifikasi.
3. Bagaimana akurasi metode WAKNN dalam proses klasifikasi yang dapat dilihat dari jumlah dokumen yang diklasifikasikan dengan benar bila dibandingkan dengan metode KNN.

Batasan masalah yang digunakan pada Tugas Akhir ini adalah :

1. Artikel berita yang digunakan bukanlah berbentuk halaman web tapi dokumen yang bersifat offline. Sumber dokumen didapat dari www.kompas.com tanggal 1 April sampai dengan 31 April 2008.
2. Artikel berita yang digunakan hanya artikel yang berbahasa Indonesia.

Hipotesis sementara adalah metode WAKNN merupakan metode yang lebih akurat untuk proses klasifikasi teks bila dibandingkan dengan metode *k-Nearest Neighbor*.

1.3 Tujuan

Berdasarkan permasalahan di atas, tujuan dari Tugas Akhir ini adalah:

1. Membangun suatu aplikasi untuk proses klasifikasi teks dengan menggunakan metode WAKNN.
2. Membandingkan akurasi proses klasifikasi teks yang dilakukan dengan metode WAKNN dengan metode klasifikasi KNN yang belum dimodifikasi.

1.4 Metodologi penyelesaian masalah

Metodologi penyelesaian yang akan dilakukan adalah:

1. Studi Literatur :
 - a. Pencarian referensi yang berhubungan dengan Tugas Akhir
 - b. Mempelajari dan lebih memperdalam materi yang berhubungan dengan Tugas Akhir.
2. Mengumpulkan artikel-artikel berita.
3. Melakukan *preprocessing* terhadap artikel-artikel berita yang telah dikumpulkan.
4. Mempelajari konsep dari metode WAKNN untuk diterapkan pada klasifikasi berita.
5. Melakukan analisis dan perancangan perangkat lunak.
6. Melakukan implementasi perangkat lunak.
7. Melakukan pengujian akurasi metode WAKNN dalam proses klasifikasi berita dengan melihat persentase dokumen yang diklasifikasikan dengan benar bila dibandingkan dengan metode klasifikasi KNN.
8. Menganalisis hasil pengujian.
9. Pengambilan kesimpulan dan penyusunan laporan Tugas Akhir.