

Abstraksi

Berkembangnya teknologi di dunia maya membuat jumlah informasi berupa artikel berita semakin banyak. Untuk itu, diperlukan suatu kategorisasi terhadap artikel yang memudahkan pembaca mencari informasi dengan menerapkan salah satu fungsionalitas dari data mining, yaitu klasifikasi. Akan tetapi, masalah utama yang terjadi pada kategorisasi artikel ini adalah tingginya dimensi dari data yang dapat mengganggu kategorisasinya itu sendiri. Oleh karena itu, harus dilakukan pemilihan terhadap beberapa atribut yang dapat berpengaruh besar terhadap hasil kategorisasi, yaitu *Feature Selection*.

Feature selection memiliki kemampuan mengurangi dimensionalitas suatu data sehingga dapat meningkatkan efektivitas dari *classifier*. Beberapa mekanisme *feature selection* telah dikembangkan dalam pengklasifikasian teks, seperti *mutual information* dan *chi-square* yang telah dianggap sebagai mekanisme yang efektif.

Pada Tugas Akhir ini, akan dibahas sebuah metode lain yang dikenal dengan nama *within class popularity* yang digunakan untuk pengklasifikasian teks. Dataset yang digunakan adalah artikel berita berbahasa Indonesia yang berasal dari web dan menggunakan beberapa *classifier*, seperti: *Naïve Bayes*, dan *k-Nearest Neighbor* (kNN).

Selain itu, dilakukan pula analisis perbandingan terhadap metode *within class popularity* dengan *mutual information* dan *chi-square* dengan menggunakan *classifier* dan berdasarkan hasil analisis diketahui bahwa *within class popularity* memiliki kinerja yang lebih baik dibanding *mutual information* dan *chi-square*.

Kata kunci : *within class popularity, feature selection, klasifikasi teks, classifier*