

Abstrak

Twitter merupakan layanan jejaring sosial dan *microblogging* yang memungkinkan pengguna untuk mengirim dan membaca pesan berbasis teks hingga 140 karakter, dengan kemajuan yang sangat pesat Twitter menjadi objek analisis yang sangat baik untuk berbagai kepentingan, salah satu penelitian yang diminati saat ini terhadap sosial media Twitter adalah analisa sentiment dan *opinion mining*. Untuk melakukan *opinion mining* terhadap Twitter menjadi kendala sendiri karena keterbatasan data Twitter dari twit pengguna yang hanya di batasi oleh 140 karakter, selain itu penelitian *opinion mining* biasanya hanya terfokus pada klasifikasi atau *clustering* data tetapi tidak banyak menjelaskan tahap *Preprocessing*, pada dasarnya *Preprocessing* yang baik akan menghasilkan proses *mining* yang baik juga, maka perlu berbagai cara untuk memaksimalkan proses *Preprocessing* pada Twitter salah satunya adalah dengan proses *stemming* dengan mengimplementasikan algoritma Soundex dimana algoritma ini diharapkan mampu memaksimalkan proses *stemming* pada *Preprocessing* untuk proses *mining* pada Twitter, selain itu metode ini akan di pasang dengan berbagai variasi algoritma pembobotan *Term Frequency (TF)*, *Feature Term Presence (TP)*, *Term Frequency-Inverse DocumentFrequency (TF-IDF)* untuk menemukan pasangan algoritma yang tepat untuk mendukung proses klasifikasi yang baik, klasifikasi dilakukan dengan metode *Naïve bayes* yang selanjutnya dapat di analisa bagaimana pengaruh algoritma soundex untuk *stemming* serta pengaruh algoritma pembobotan jika diterapkan pada proses klasifikasi, selain itu penelitian diharapkan mampu menghasilkan algoritma yang memberikan kontribusi yang baik untuk proses *stemming* data Twitter serta mempelajari bagaimana pengaruh algoritma pembobotan jika dipasangkan dengan algoritma soundex. Setelah dilakukan penelitian terhadap *stemming* dibandingkan hasil *stemming* algoritma soundex dengan porter maka didapatkan hasil untuk data uji sebanyak 300 twit bahwa soundex sedikit lebih unggul kemudian diklasifikasikan data hasil *stemming* dengan soundex dengan beberapa algoritma pembobotan didapatkan hasil nilai akurasi yang sama, berdasarkan analisis didapatkan bahwa algoritma pembobotan tidak berpengaruh kepada hasil klasifikasi.

Kata kunci : Twitter, *Algoritma Soundex*, *Stemming*, *Preprocessing*, *Naïve bayes*