

Pada penelitian ini akan dilakukan identifikasi parafrasa pada dua buah kalimat yang berbasis *user generated* dimana semua data murni diambil dari pengguna jejaring sosial *Twitter* yaitu *tweet* atau kicauan terbatas 140 karakter dengan menggunakan algoritma pencarian jarak antara dua buah *string* yaitu algoritma *Levenshtein Distance* dan juga akan diukur tingkat akurasi algoritma tersebut menggunakan *confusion matriks* dan juga skema pengujian menggunakan *K-Fold Cross Validation*. Dari hasil penelitian ini diperoleh nilai akurasi sebesar 0,7695, f1 sebesar 0,6906, *precision* sebesar 0,8090, dan juga nilai *recall* sebesar 0,6024.