**Filtering dan Klasifikasi Data**

Setelah pada tahap sebelumnya anda telah melakukan proses *scraping* dan mendapatkan hasil *scrap* dalam bentu csv file, selanjutnya anda dapat melakukan proses *filtering* pada data csv yang ada dengan menekan tombol bertuliskan **NAMA\_TOMBOL** tersebut. Seperti pada gambar di bawah ini.



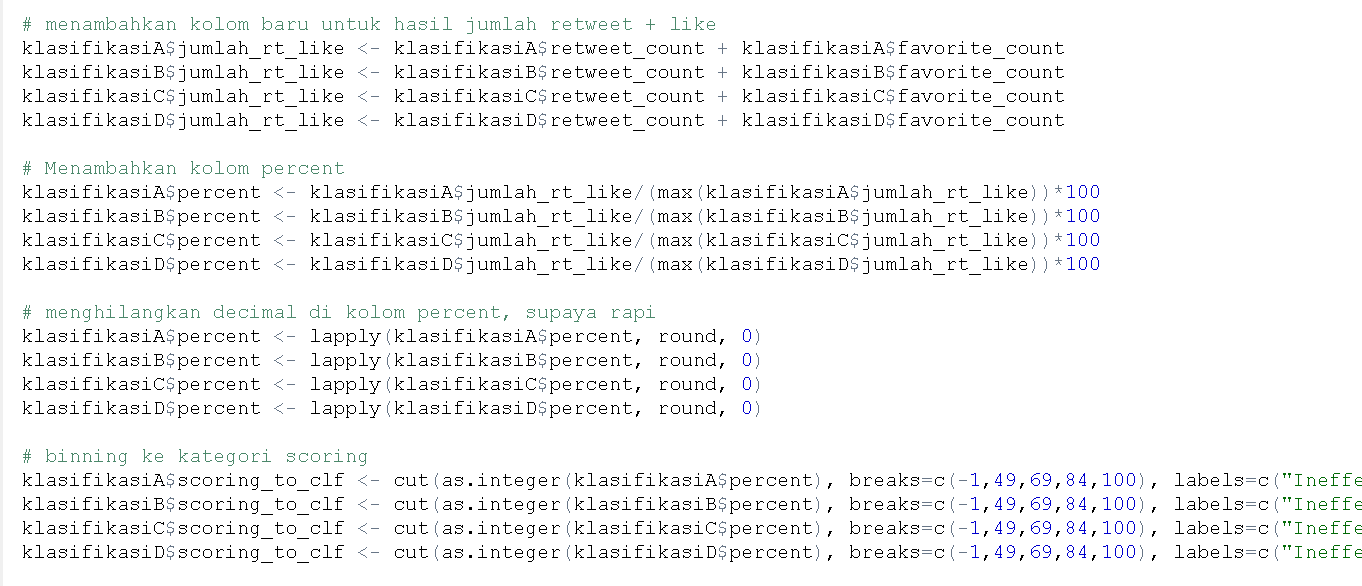
**Gambar 1.1** *Tampilan User Interface*

Pada saat anda menekan *button* tersebut, maka proses yang terjadi adalah, program akan menjalankan serangkaian fungsi yang akan menghasilkan data baru yang nantinya akan siap digunakan sebagai data dalam proses *scalling*, pada tahap selanjutnya. Pertama, program akan melakukan *import datasets* dari data hasil *scraping* yang sudah dilakukan. Selanjtunya merubah semua data yang ada di dalam kolom text menjadi *lowercase*, atau huruf kecil. Kemudian, setelah semua data di dalam kolom text menjadi *lowercase*, proses *filtering* pertama dilakukan, yaitu melakukan filter untuk mendaaptkan data yang hanya menggunakan bahasa English dan Indonesia. Setelah filter tersebut dilakukan maka data yang tersisa adalah data yang menggunakan kedua bahasa tersebut. Kemudian proses *filtering* selanjutnya adalah melakukan filter berdasarkan klasifikasi yang sudah ditentukan pada ketentuan awal. Klasifikasi dilakukan dengan cara mencari data yang memiliki kata sesuai dengan *keyword* awal yang sudah ditentukan. Proses klasifikasi data dilihat pada gambar di bawah ini.



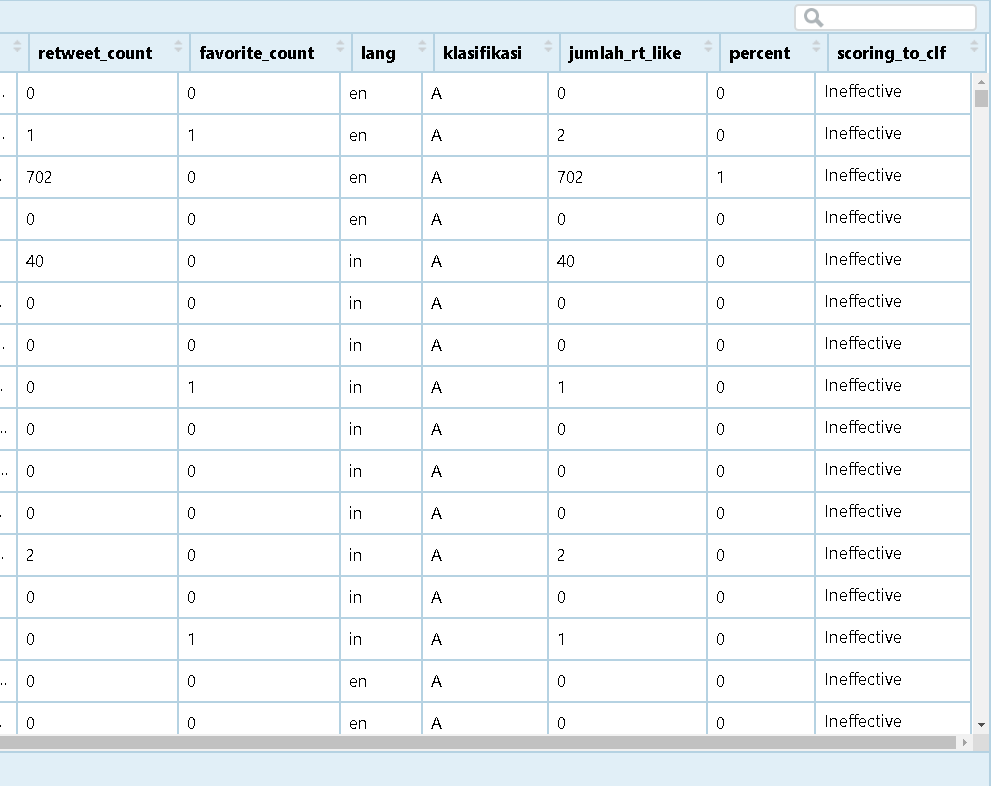
**Gambar 1.2** *Sintaks klasifikasi*

Kemudian, setelah proses klasifikasi dilakukan, selanjutnya dilakukan proses perhitungan jumlah retweet dan jumlah like. Masing-masing dari kalsifikasi tersbeut, akan dibuat kolom baru yang berisi hasil penjumlahan dari retweet dan like. Kemudian dilakukan perhitungan dari tiap jumlah tersebut berapa persen terhadap nilai maksimal yang ada. Lalu, hasil dari kolom percent tersebut akan dilakukan binning untuk mengelompokkan nilai-nilai sesuai dengan kategori penilaian, seperti efektif, moderately, dll.



**Gambar 1.3** *Sintaks klasifikasi dan binning*

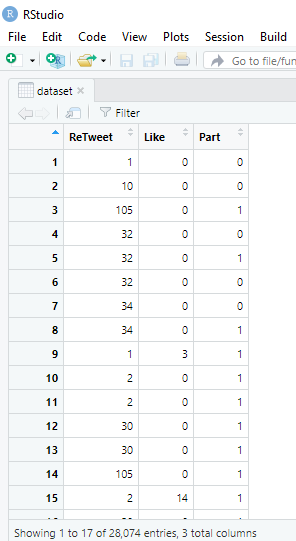
Hasil dari semua klasifikasi, perhitungan jumlah retweet dan like, hingga binning kategori peniliaian akan digabungkan ke dalam satu file csv yang baru, yang nantinya akan menjadi datasets di tahapan selanjutnya. Hasil dari tahapan filtering dan klasifikasi adalah sebagai berikut.



**Gambar 1.4** *Hasil klasifikasi*

**Scalling**

Setelah semua data diproses dalam tahap sebelumnya, tahap selanjtunya adalah proses *scalling*. Pada tahap ini, anda dapat menekan *button* dengan nama **NAMA\_TOMBOL**. Dengan menekan tombol tersebut, makan program akan menjalankan serangkaian proses yang dilakukan untuk *scalling*, seperti mengganti label kategori yang dibuat dalam tahap selanjutnya menjadi label numerik, dan hanya dipilih yang memiliki label 0 dan 1. Hasil dari *scalling* dapat dilihat pada gambar berikut.

**

**Gambar 1.4** *Hasil scalling*