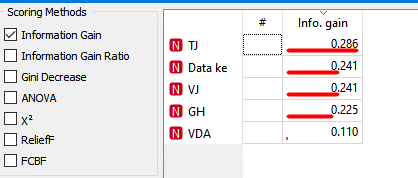
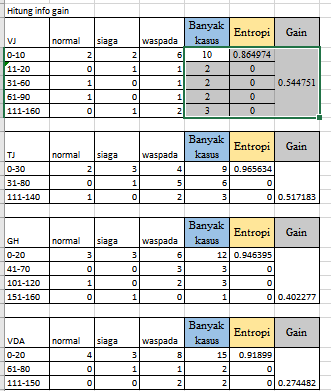
UAS Pengenalan Pola

Dyline Melynea F/185314125

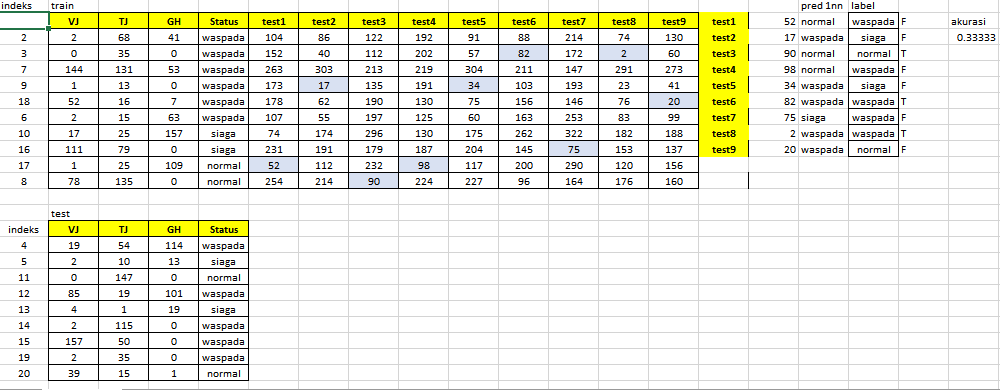
Data Gunung Merapi terdapat 19 entri. Dari 19 tersebut, terdapat 6 kolom yang 1 kolomnya menjadi label yaitu kolom ‘status’. Kolom lainnya selain ‘status’ dianggap sebagai fitur kecuali kolom ‘Data ke’ karena kolom tersebut berfungsi sebagai index. Maka didapatkan 4 feature dan 1 label. Namun, dilakukan juga proses seleksi ciri untuk menentukan feature/ciri mana yang memberikan dampak.

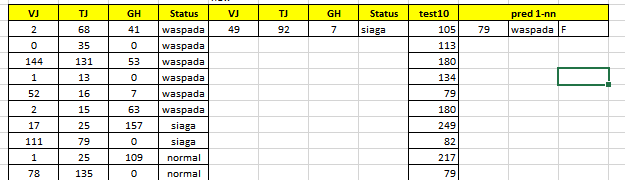
Berikut adalah perhitungan untuk menyeleksi ciri dengan software Orange:

Selain itu juga dilakukan perhitungan manual di excel yang sebelumnya membuat kelompok/range dari masing-masing feature agar membentuk kelasnya (detail ada di file data sheet ‘f.selection’):

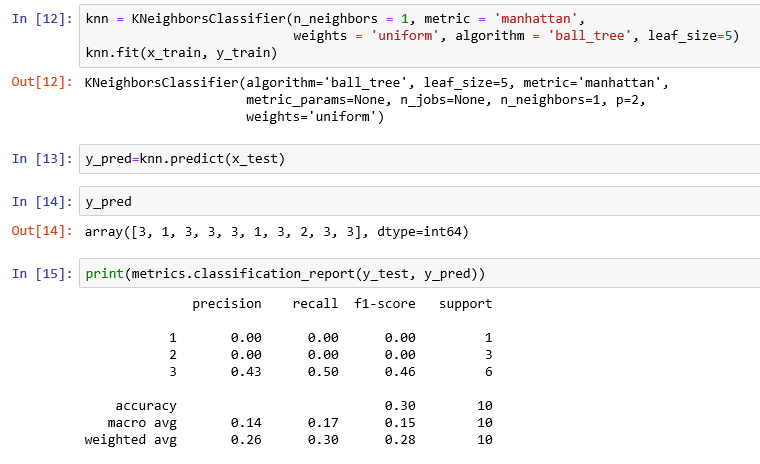
Dari kedua cara tersebut, sama-sama dihasilkan fitur yang layak digunakan ada 3 yaitu: ‘VJ’, ‘TJ’, dan ‘GH’

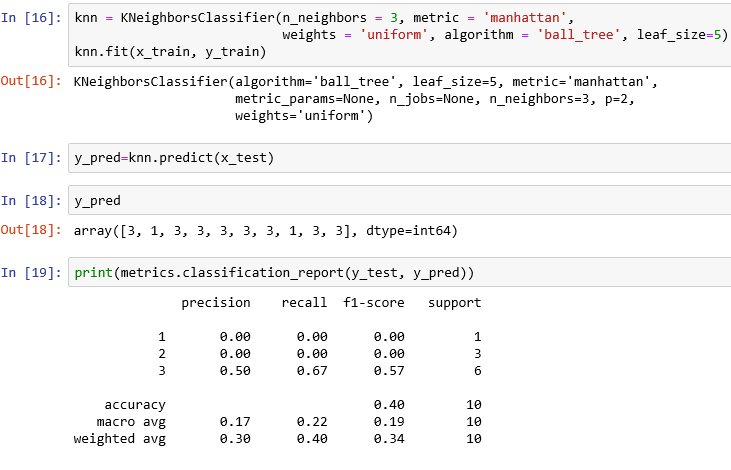
Kemudian dilakukan klasifikasi. Pertama dicoba dengan manual menggunakan excel, data dibagi menjadi 10 untuk training, dan 9 untuk testing. Digunakan KNN dengan metric/distance Manhattan dan banyak neighbor yaitu 1 untuk manual.

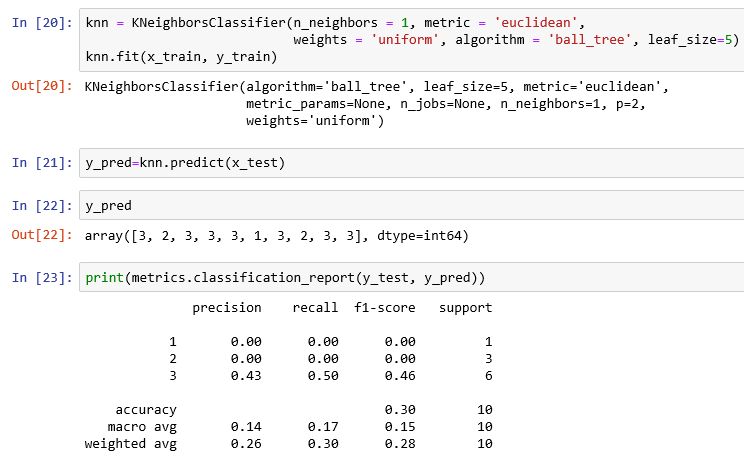
Didapatkan hasilnya seperti berikut (detail ada di file data sheet ‘data olahan’):

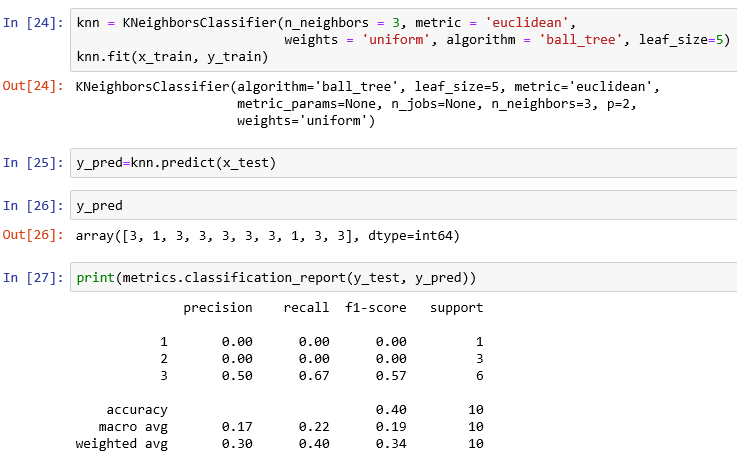
Dicoba juga ketika data baru masuk yaitu mendapatkan status waspada, dan hasilnya adalah tidak sesuai dengan kelas labelnya.

Dicoba juga dengan program menggunakan bahasa python3. Data dibagi menjadi 50% training, dan 50% testing. Berikut adalah dokumentasi dan rekapan dari percobaan tuning parameter.

**Manhhatan, neighbot = 1, akurasi= 30%**

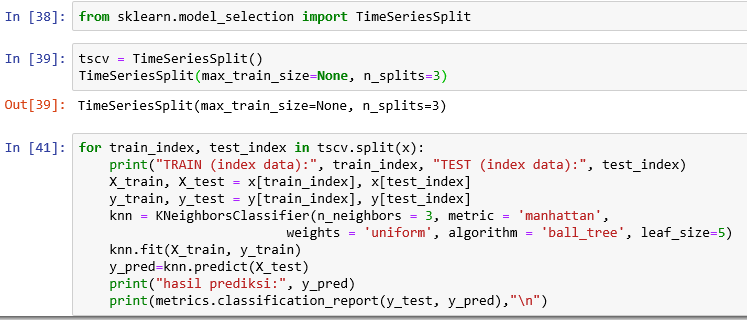
**Manhattan, neighbors = 3, akurasi 40%**

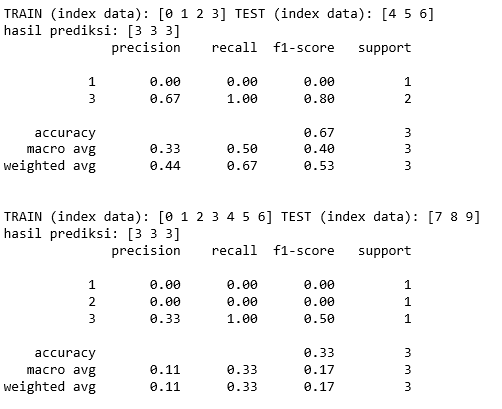
**Euclidean, neighbors = 1, akurasi = 30%**

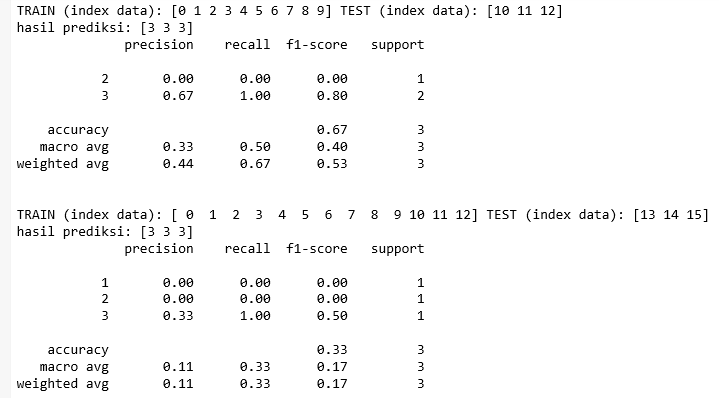
**Euclidean, neighbors = 3, akurasi = 40%**

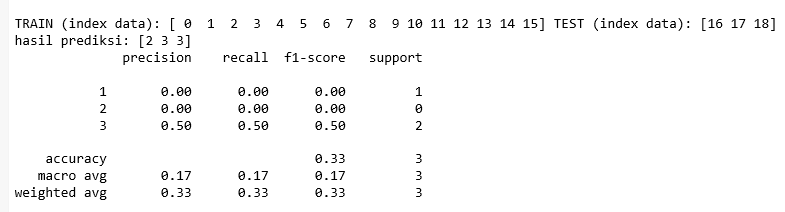
Digunakan maksimal banyak neighborsnya 3 karena ada 2 kelas yang memiliki entri hanya 4, dan jumlah neighbor lebih baik ganjil, sehingga dipilih 1 atau 3, dari hasil program ini maka didapatkan hasil paling baik yaitu dengan neighbors sebanyak 3 dengan akurasi 40% baik untuk metric/distance Manhattan maupun Euclidean.

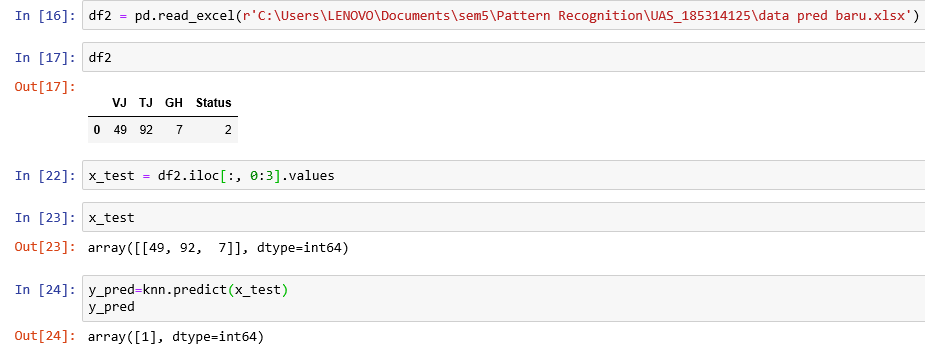
Kemudian juga dilakukan Time series cross validation yaitu seperti berikut:

**n\_split = 3**





Dari percobaan time series cross validation tersebut didapatkan hasil yang paling baik yaitu dengan skor 67% untuk iterasi pertama dan ketiga, detail indeks data yang dipakai ada seperti gambar mengenai hasil cross validation di atas. Untuk knn, konfigurasinya yang digunakan yaitu dengan Manhattan dan banyak neighbors = 3 karena neighbors 3 adalah yang paling baik sebelumnya, dan digunakan Manhattan karena hasil dari Euclidean maupun Manhattan adalah sama sebelumnya, sehingga dipilih Manhattan agar lebih sesuai dengan percobaan manual menggunakan excel.

Dari program juga dicoba untuk memprediksi data baru, didapatkan hasilnya yaitu seperti berikut:

Hasil prediksi terhadap data baru yaitu ‘**1’** yang artinya statusnya normal. Hasilnya tidak sesuai dengan yang dicantumkan di kelas label.

Kesimpulan:

Dengan rangkaian percobaan di atas, didapatkan hasil akurasi maksimal yaitu 67%. Dan juga setelah dilakukan percobaan pada data baru, hasilnya baik menggunakan excel maupun program yaitu sama-sama tidak sesuai dengan kelas labelnya. Dapat dikatakan bahwa model KNN kurang cocok untuk prediksi data Gunung Merapi.