

Tugas 4
CSH3L3 Pembelajaran Mesin Genap 2018/2019

Nama : Fikri Rozan Imadudin/1301150768

Kelas : IF-40-10

Analisis dan strategi penyelesaian masalah.

Tujuan dan Masalah :

1. Diberikan data TrainsetTugas4ML.csv dan TestsetTugas4ML.csv yang memiliki label yaitu X1,X2 dan class. TrainsetTugas4ML akan digunakan sebagai data training dan TestsetTugas4ML akan digunakan sebagai data testing.
2. Pada tugas 4 kali ini akan dibuat sebuah sistem klasifikasi menggunakan metode metode *Bagging*, salah satu teknik *Ensemble Learning*, berbasis *Naïve Bayes* untuk menentukan kelas/label data uji dalam TestsetTugas4ML.

Teori dan Kajian :

Bagging merupakan salah satu dari *Ensemble Learning* yang digunakan untuk mengurangi variansi, meningkatkan akurasi algoritma klasifikasi dan dapat menghindari *overfitting*. Idennya yaitu membagi data kedalam beberapa *subspace* lalu mengkombinasikan data secara acak dengan ukuran yang sama dengan data asli dalam masing-masing subspace sehingga didapatkan dari masing-masing *subspace* rata-rata untuk prediksi atau klasifikasi yang akan digunakan nanti.

Metode Penelitian :

Metodologi yang digunakan terdiri dari beberapa tahap yaitu memasukan dataset, memisahkan dataset menjadi dua bagian yaitu x(X1,X2), y(class), perancangan sistem, pengujian model dan analisa hasil. Akan digunakan sebuah library sklearn BaggingClassifier yang merupakan metode bagging dan from sklearn GaussianNB sebagai naive bayes. Lalu digunakan KFold untuk mengevaluasi model.

Analisis dan Hasil :

Experiment dilakukan terhadap *Naïve Bayes* tanpa menggunakan *Bagging* dan *Naïve Bayes* dengan *Bagging*. Pada tabel berikut berupa hasil yang telah diterapkan dengan `random_state = 7` dan `n_estimator pada bagging = 100`.

Algoritma	Akurasi	KFold
<i>Naïve Bayes</i> tanpa menggunakan <i>Bagging</i>	93.9597 %	87.66 %
<i>Naïve Bayes</i> dengan <i>Bagging</i>	94.6308 %	86.99 %

Hasil experiment yang ada pada tabel berikut menyatakan bahwa penerapan teknik bagging pada *Naïve Bayes* meningkatkan nilai akurasi sebesar 0.6711 % dan hasil KFoldnya *Naïve Bayes* lebih unggul. Pada percobaan ke dua kali ini `random_state = 10` dan `n_estimator pada bagging = 100`.

Algoritma	Akurasi	KFold
<i>Naïve Bayes</i> tanpa menggunakan <i>Bagging</i>	93.9597 %	87.6 %
<i>Naïve Bayes</i> dengan <i>Bagging</i>	93.9597 %	86.9 %

Hasil experiment yang ada pada tabel berikut menyatakan bahwa penerapan teknik *Naïve Bayes* tanpa menggunakan *Bagging* hasil Kfoldnya lebih tinggi dan akurasinya sama. Pada percobaan ke tiga kali ini `random_state = 7` dan `n_estimator pada bagging = 9`.

Algoritma	Akurasi	KFold
<i>Naïve Bayes</i> tanpa menggunakan <i>Bagging</i>	93.9597 %	87.6 %
<i>Naïve Bayes</i> dengan <i>Bagging</i>	93.9597 %	88 %

Hasil experiment yang ada pada tabel berikut menyatakan bahwa penerapan teknik *Naïve Bayes* menggunakan *Bagging* hasil Kfoldnya lebih tinggi dan akurasinya sama. Pada percobaan ke empat kali ini `random_state = 10` dan `n_estimator` pada *bagging* = 9.

Algoritma	Akurasi	KFold
<i>Naïve Bayes</i> tanpa menggunakan <i>Bagging</i>	93.9597 %	87.66 %
<i>Naïve Bayes</i> dengan <i>Bagging</i>	94.6308 %	87.66 %

Hasil experiment yang ada pada tabel berikut menyatakan bahwa penerapan teknik *Naïve Bayes* menggunakan *Bagging* hasil lebih tinggi akurasinya dan KFoldnya sama.

Kesimpulan :

Pada percobaan satu dengan teknik *bagging* Akurasinya unggul tetapi KFold lebih rendah, lalu pada percobaan ke empat akurasi *bagging* lebih unggul dan KFoldnya sama. Teknik *Bagging* terbukti meningkatkan akurasi walaupun tidak besar karena data yang digunakan jumlahnya sedikit dan dari experiment tersebut bahwa `subspace/n_estimatornya` dan `random_state` mempengaruhi nilai dari akurasi dan KFold.