

Nama : : Endang Prayoga Hidayatulloh
NIM : : 2006109
Mata Kuliah : Machine Learning

Ensemble Learning

Ensemble Learning merupakan metode gabungan dengan salah satu tujuannya dapat meningkatkan kinerja suatu algoritma. Berikut adalah konsep dari ensemble learning.

7 Konsep Dasar Ensemble Learning

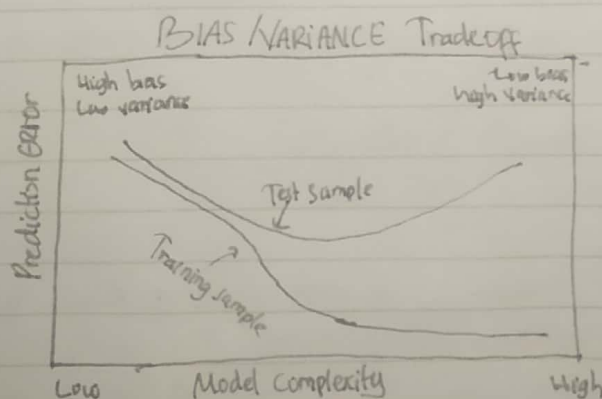
Sebagai contoh dari kasus seleksi model yang di hasilkan pada tabel dibawah, ditanyakan manakah model terbaik yang di hasilkan dari 4 algoritma berbeda?

Classifier	Accuracy
KNN	73%
Decision Tree	79%
Naive Bayes	75%
CART	76%

- Dari hasil tersebut dihasilkan model classifier dengan akurasi terbaik yaitu Decision Tree.
- Tetapi kemungkinan sayangnya, model algoritma lain akan dihapus ataupun di buang dari hasil seleksi tersebut.

- Dan hasil masing-masing model juga memiliki kemampuannya untuk mempelajari properti spesifik yang melekat didalam kumpulan data.

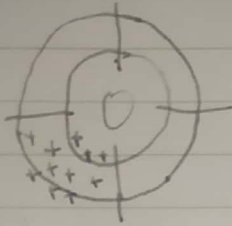
Sebelum menuju ke ensemble learning, kita terlebih dahulu harus tahu bagaimana konsep algoritma tunggal.



- Bias, Perbedaan prediksi dari nilai sebenarnya
high bias : tidak beradaptasi dengan baik
- Variance, Perubahan model karena set pelatihan yang berbeda.
high variance : Sangat cocok dengan set pelatihan.

Beberapa kondisi bias dan Variance

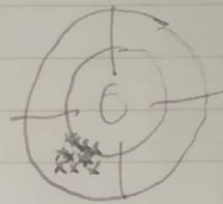
> High Bias dan High Variance



- Kondisi model terburuk
- Prediksi error sangat tinggi
- Prediksi sensitif terhadap perubahan data pelatihan

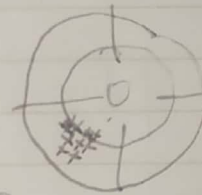
> High Bias dan Low Variance

- Underfit dari sebuah model
- Model yang dihasilkan ketika dataset kurang (terlalu sedikit)



> Low Bias dan High Variance

- Overfitting dari sebuah model
- Kondisi ini dihasilkan saat dataset mengandung banyak noise.
- Algoritma yang rentan overfitting diantaranya Decision Tree, SVM, Neural Network.



> Low Bias dan Low Variance



- Kondisi Ideal dari sebuah model
- Kondisi error kemungkinan kecil
- Prediksi tidak boleh berubah karena perubahan data pelatihan

Untuk mencapai kondisi yang ideal ini sebagai berikut:

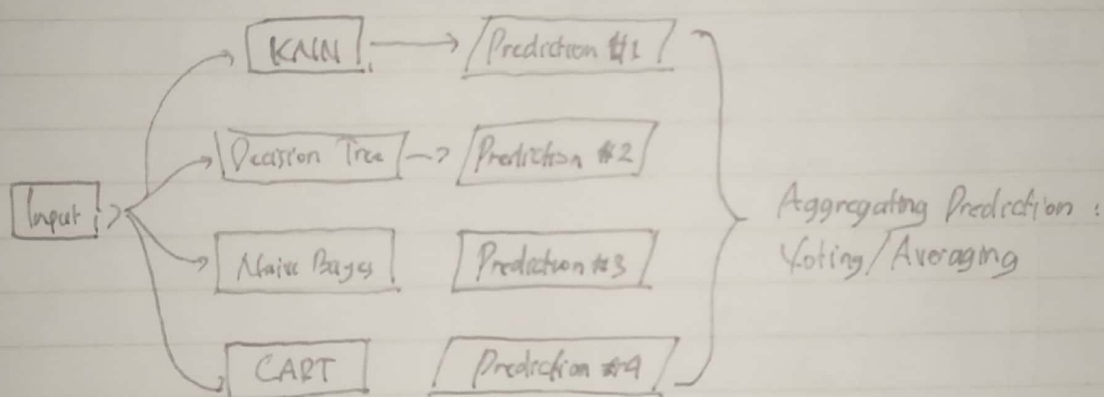
- Dataset yang digunakan harus bisa diandalkan, ketika membuat ataupun mengumpulkannya.
- Cross Validation, meminimalkan overfitting.
- Best-fit Model Parameter
- Hyperparameter Tuning
- Pemilihan metrik model evaluasi : Accuracy, Precision, Recall, AUC, F1-Score
- Melakukan Komparasi Algoritman
- Melakukan Ensemble Learning.

Ensemble Learning terinspirasi dari sebuah kondisi ketika pengambilan keputusan, dibuat keputusan yang diambil dari suatu sisi, hasilnya akan bersifat subjektif karena sudut pandangnya hanya diambil dari satu sisi saja. Dan kita tidak akan tahu bagaimana pendapat dari sisi yang lain. Oleh karena itu kita bisa melakukannya secara kolektif untuk mengatasi hal itu.

Kombinasi Model \rightarrow Ensemble Learning

Dari hasil penggabungan model didapatkan

Combined Classifier	Accuracy
KNN + Decision Tree + Naive Bayes + CART	85%



\therefore Dengan menjumlahkan (voting) keputusan dari classifier tersebut dapat meningkatkan akurasi atau setidaknya sama dengan classifier tunggal terbaik.

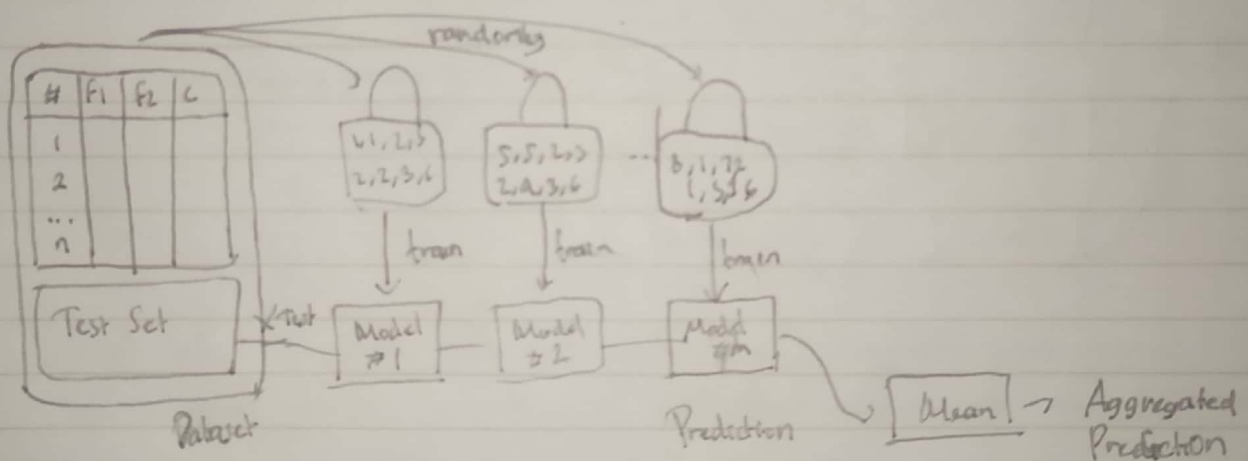
Terdapat 3 Model Ensemble Methods; diantaranya.

- Bagging
- Random Forest
- Boosting

> Bagging Method

Bagging adalah singkatan dari Bootstrap Aggregating. Menupakan teknik membuat ensemble dengan berulang kali melakukan resampling data pelatihan secara acak. Berikut adalah algoritmanya:

1. Diketahui bahwa D sebagai dataset asli dengan ukuran n .
2. lalu setelah D menjadi dataset, buat m bootstrap sample dengan ukuran n^* dengan menggambarkan n^* contoh dari data asli. Dengan penggantian.
 - $n^* < n$; $n^* \sim 60\% \cdot n$
 - Setiap bootstrap sample rata-rata akan berisi 63,2% dari contoh pelatihan unik, sisanya direplikasi.
3. Menggabungkan m model yang dihasilkan menggunakan suara mayoritas sederhana.
4. Mengurangi kesalahan dengan mengurangi varian dalam hasil karena training yang tidak stabil, algoritma (seperti decision tree) yang keduanya dapat berubah secara dramatis saat data pelatihan sedikit.



Kesimpulannya:

Teknik ini mereduksi variansi atau overfitting dengan melakukan rata-rata keputusan. Hasilnya mereduksi MSE namun hampir tidak menurunkan bias atau bahkan tidak mempengaruhi sama sekali.

Lalu bagaimana seandainya kita mau menurunkan bias tanpa mengganggu variansi? Di sini kita bisa menggunakan teknik yang bernama boosting

> Boosting

Secara konsep boosting adalah teknik untuk memperkuat pelatihan model dengan cara mengulangi pelatihan untuk data-data yang masih salah terklasifikasi. Salah satu metode boosting yang populer adalah AdaBoost.

AdaBoost, algoritma ini memberikan bobot pada setiap data sebagai tingkat probabilitas seberapa mungkin data tersebut diambil untuk dilatih, data yang salah memprediksi akan dinaikkan bobotnya sementara data yang sudah benar akan diturunkan bobotnya. Hasil prediksi akhir adalah weighted average semua model yang terbentuk berdasarkan bobot data latih yang digunakan.

Kesimpulannya.

Kelebihan boosting adalah implementasi sederhana dalam mereduksi variansi error. Kekurangannya teknik ini cukup sensitif terhadap noise data yang outlier.