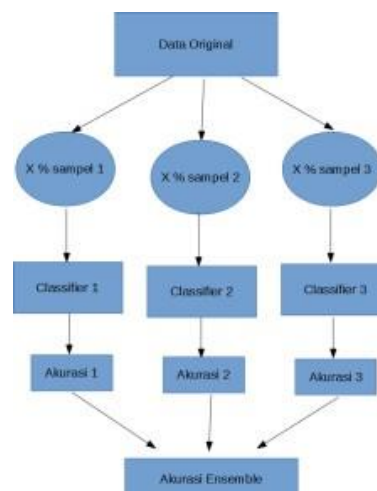


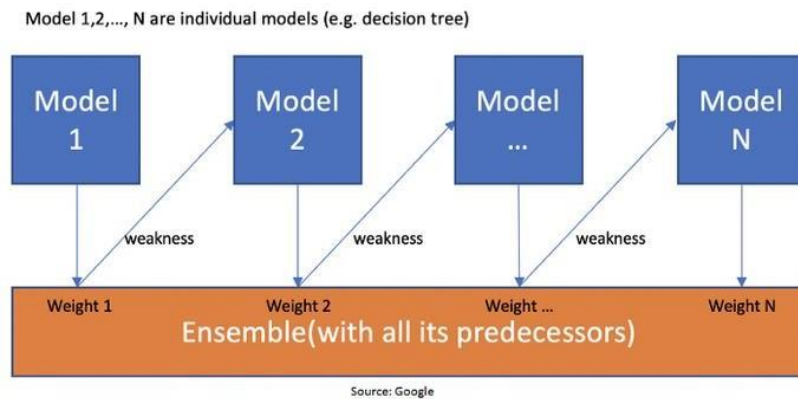
1. Ensemble adalah Teknik untuk menggabungkan beberapa metode dalam melakukan klasifikasi pada data sehingga dalam metode ensemble terdapat N buah classifier atau disebut base estimator, N buah classifier ini berjenis sama, misalnya KNN semua, Naïve bayes semua atau SVM semua.
2. Perbedaan dari Ensemble Bagging, Boosting dan Stacking :
 - Bagging adalah metode pembangkitan ensemble yang menggunakan variasi sampel yang digunakan untuk melatih pengklasifikasi dasar. Untuk setiap pengklasifikasi yang akan dihasilkan, Bagging memilih dengan pengulangan N sampel dari set pelatihan dengan ukuran N dan melatih pengklasifikasi dasar. Ini diulang sampai ukuran ansambel yang diinginkan tercapai.
 - Boosting pendekatan yang menarik tetapi sangat sensitif terhadap noise dan hanya efektif menggunakan pengklasifikasi yang lemah. Ada beberapa variasi teknik Boosting AdaBoost, BrownBoost (...), masing-masing memiliki aturan pembaruan bobot sendiri untuk menghindari beberapa masalah tertentu (noise, ketidakseimbangan kelas...).
 - Stacking adalah pendekatan meta learning dimana ensemble digunakan untuk "mengekstrak fitur" yang akan digunakan oleh lapisan ensemble lain.
3. Konsep kerja dari Algoritma Ensemble Bagging, Boosting dan Stacking dan ilustrasikan dengan gambar :
 - Bagging

Konsep kerja : pada bagging sejumlah classifier dibangun dengan masing-masing menggunakan x% dari data. Masing-masing classifier akan melakukan klasifikasi pada sebanyak x% dari data yang diambil secara random. Masing- masing classifier ini akan menghasilkan akurasi dari ensemble bagging secara keseluruhan adalah rata-rata dari akurasi yang dihasilkan masing-masing classifier.



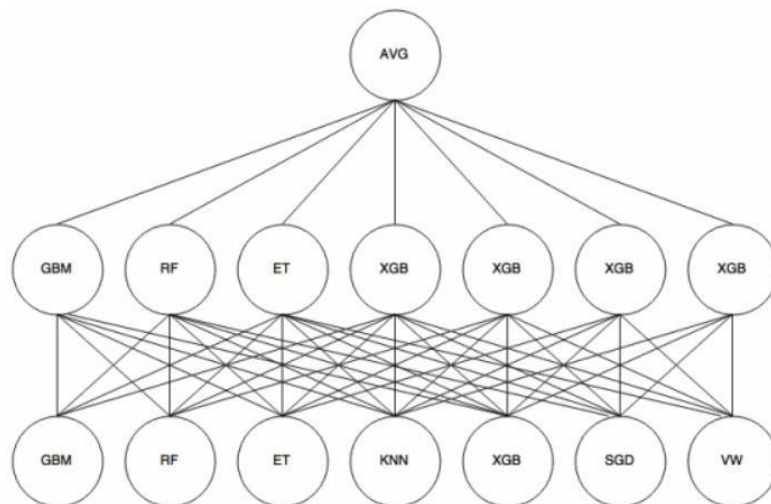
- Boosting

Konsep kerja : Ini membuat sejumlah pohon keputusan saat kita melatih data. Saat pohon keputusan atau model pertama dibuat, contoh pelatihan salah mengklasifikasikan mode pertama, lalu model pertama memiliki prioritas lebih. Hanya contoh pelatihan ini yang dikirim sebagai masukan untuk model kedua. Gambar di bawah menunjukkan bahwa ketika model pertama dibuat, dan algoritme memperhatikan kesalahan model pertama, contoh pelatihan yang salah diklasifikasikan diberikan sebagai masukan untuk model berikutnya. Proses ini diulang terus menerus hingga kondisi yang ditentukan tidak terpenuhi. Jika Anda melihat gambar di atas, ada n jumlah model yang dibuat dengan meminimalkan kesalahan model sebelumnya. Beginilah cara kerja penguat. Model 1,2, 3,..., N adalah model individu yang dikenal sebagai pohon keputusan. Semua metode ensembling bekerja pada konsep yang sama.



- Stacking

Konsep kerja : Pertama (Bawah) beberapa pengklasifikasi berbeda dilatih dengan set pelatihan, dan keluarannya (probabilitas) digunakan untuk melatih lapisan berikutnya (lapisan tengah), akhirnya, keluaran (probabilitas) pengklasifikasi di lapisan kedua digabungkan menggunakan rata-rata (AVG). Ada beberapa strategi yang menggunakan validasi silang, pencampuran, dan pendekatan lain untuk menghindari penumpukan yang berlebihan. Tetapi beberapa aturan umum adalah untuk menghindari pendekatan seperti itu pada kumpulan data kecil dan mencoba menggunakan pengklasifikasi yang beragam sehingga mereka dapat "melengkapi" satu sama lain.

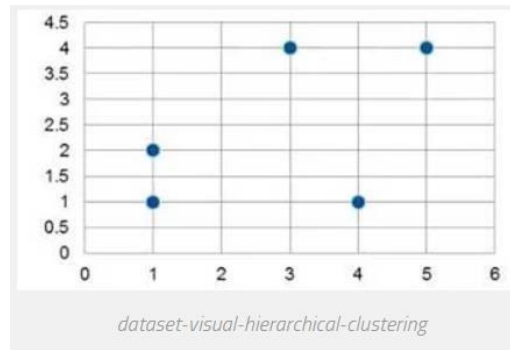


4. Contoh Kasus dan penyelesaian menggunakan Algoritma Hierarchical Clustering disertai dengan Langkah – Langkah penyelesaiannya:

Data set :

Data	Fitur x	Fitur y
1	1	1
2	4	1
3	1	2
4	3	4
5	5	4

Kelompokkan dataset dengan menggunakan algoritma hierarchical clustering (single linkage) menggunakan jarak Manhattan



Mengitungh jarak pada semua pasangan dua data denga persamaan berikut :

$$D = \sum_{i=1}^n |b_i - a_i|$$

$D_{man}(Data1, Data1) = |1-1| + |1-1| = 0$
 $D_{man}(Data1, Data2) = |1-4| + |1-1| = 3$
 $D_{man}(Data1, Data3) = |1-1| + |1-2| = 1$
 $D_{man}(Data1, Data4) = |1-3| + |1-4| = 5$
 $D_{man}(Data1, Data5) = |1-5| + |1-4| = 7$
 $D_{man}(Data2, Data3) = |4-1| + |1-2| = 4$
 $D_{man}(Data2, Data4) = |4-3| + |1-4| = 4$
 $D_{man}(Data2, Data5) = |4-5| + |1-4| = 4$
 $D_{man}(Data3, Data4) = |1-3| + |2-4| = 4$
 $D_{man}(Data3, Data5) = |1-5| + |2-4| = 6$
 $D_{man}(Data4, Data5) = |3-5| + |4-4| = 2$

2 Dengan table :

Dman	1	2	3	4	5
1	0	3	1	5	7
2	3	0	4	4	4
3	1	4	0	4	6
4	5	4	4	0	2
5	7	4	6	2	0

Cara Metode single linkage :

Dengan memperlakukan data sebagai kelompok, selanjutnya kita pilih jarak dua kelompok yang terkecil.

$\min(D_{man}) = \min(d_{13}) = 1$

Terpilih kelompok 1 dan 3, sehingga kedua kelompok ini digabungkan.

Menghitung jarak antar kelompok (1 dan 3) dengan kelompok lain yang tersisa, yaitu 2, 4 dan 5.

$d_{(13)2} = \min \{d_{12}, d_{32}\} = \min \{3, 4\} = 3$

$d_{(13)4} = \min \{d_{14}, d_{34}\} = \min \{5, 4\} = 4$

$d_{(13)5} = \min \{d_{15}, d_{35}\} = \min \{7, 6\} = 6$

Dengan menghapus baris-baris dan kolom-kolom matrik jarak yang bersesuaian dengan kelompok 1 dan 3, serta menambahkan baris dan kolom untuk kelompok (13).

Dman	(13)	2	3	4
(13)	0	3	4	6
2	3	0	4	4
4	4	4	0	2
5	6	4	2	0

Selanjutnya dipilih jarak dua kelompok yang terkecil.

$\min(D_{man}) = \min(d_{45}) = 2$

Menghitung jarak antar kelompok (4 dan 5) dengan kelompok lain yang tersisa, yaitu (13) dan 2.

$$d(45)(13) = \min \{d_{41}, d_{43}, d_{51}, d_{53}\} = \min \{5, 4, 7, 6\} = 4$$

$$d(45)2 = \min \{d_{42}, d_{52}\} = \min \{4, 4\} = 4$$

Menghapus baris dan kolom matrik yang bersesuaian dengan kelompok 4 dan 5, serta menambahkan baris dan kolom untuk kelompok (45)

Dman	(45)	(13)	2
(45)	0	4	4
(13)	4	0	3
2	4	3	0

Selanjutnya dipilih jarak dua kelompok yang terkecil.

$$\min(D_{man}) = \min(d(13)2) = 3$$

Terpilih kelompok (13) dan 2, sehingga kedua kelompok ini digabungkan. (Melanjutkan pengelompokan).

Menghitung jarak antar kelompok ((13) dan 2) dengan kelompok lain yang tersisa, yaitu (45). $d(132)(45) = \min \{d_{14}, d_{15}, d_{34}, d_{35}, d_{24}, d_{25}\} = \min \{5, 7, 4, 6, 4, 4\} = 4$

Menghapus baris dan kolom matrik yang bersesuaian dengan kelompok (13) dan 2, serta menambahkan baris dan kolom untuk kelompok (123).

Dman	(123)	(45)
(123)	0	4
(45)	4	0

Jadi kelompok (132) dan (45) digabung untuk menjadi kelompok tunggal dari lima data, yaitu kelompok (13245) dengan jarak terdekat 4.

Berikut Dendrogram Hasil Metode Single Linkage :

