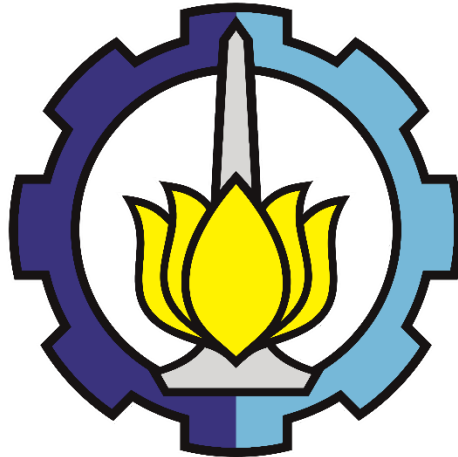


DECISION TREE & SVM
KECERDASAN KOMPUTASIONAL - B



Dosen : Ibu Dr. Diana Purwitasari, S.Kom., M.Sc.

Kelompok 3

Kevin Angga Wijaya	05111840000024
Angelita Titiandes Br. Silalahi	05111840000088
Aflakah Nur Farhana	05111840000120

DEPARTEMEN TEKNIK INFORMATIKA
FAKULTAS TEKNOLOGI ELEKTRO DAN INFORMATIKA CERDAS
INSTITUT TEKNOLOGI SEPULUH NOPEMBER SURABAYA

2020

Berikut merupakan hasil percobaan penggunaan Decision Tree dan SVM. Data yang digunakan adalah *Car Evaluation* dataset sebanyak 1727 data. Dilakukan perubahan data *categorical* kedalam bentuk *numeric* dengan keterangan sebagai berikut :

Class values :	Maint :	Buying :
Unacc = 1	Low = 1	Low = 1
Acc = 2	Med = 2	Med = 2
Good = 3	High = 3	High = 3
Vgood = 4	Vhigh = 4	Vhigh = 4

Lug_boot :	Safety :	Person :
Small = 1	Low = 1	1 = 1
Med = 2	Med = 2	2 = 2
Big = 3	High = 3	3 = 3
		4 = 4
		More = 5

Doors :
1 = 1
2 = 2
3 = 3
4 = 4
5More = 5

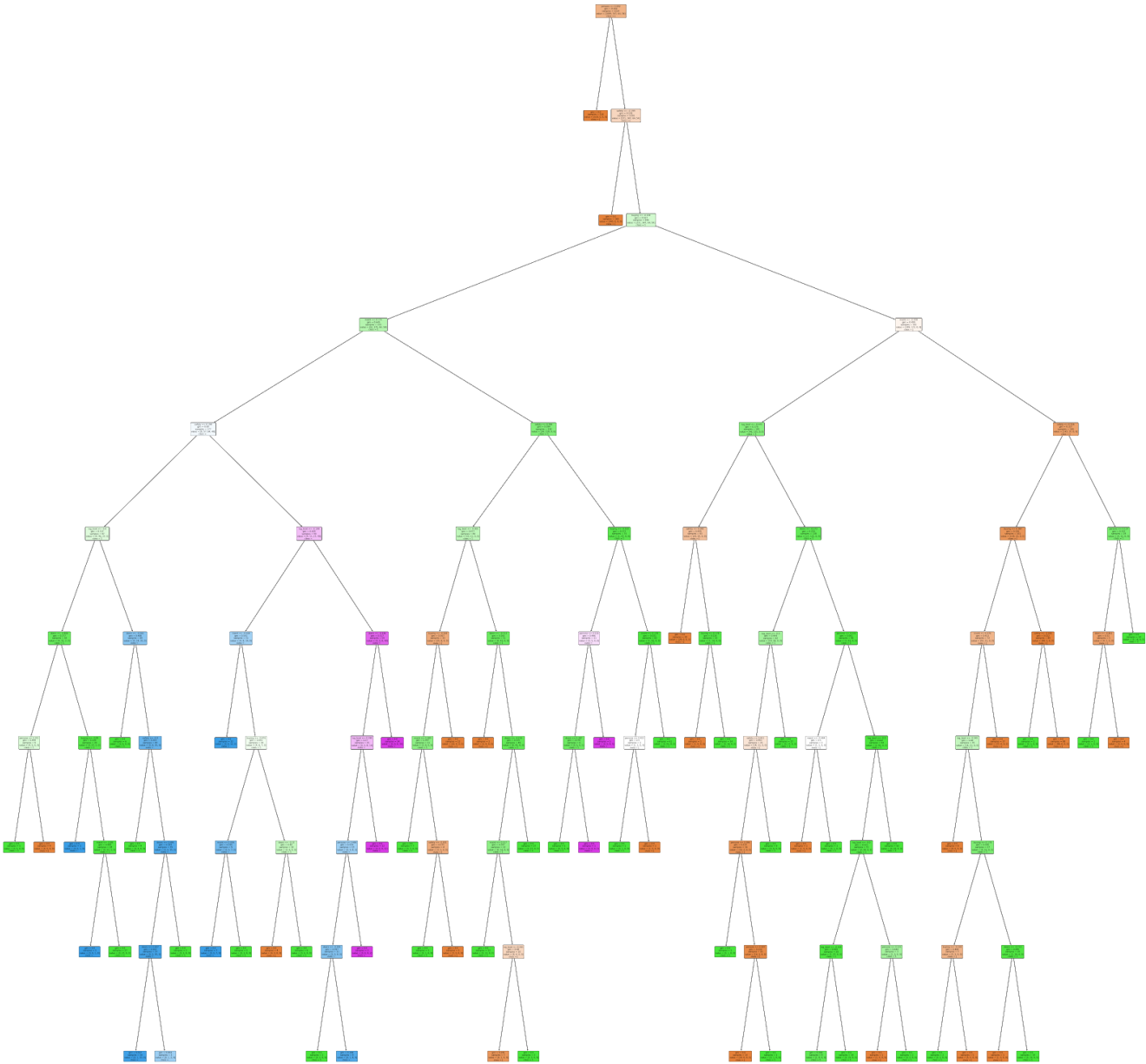
Decision Tree

Berikut merupakan hasil dari decision tree dengan melakukan variasi pada depth nya

	ACC/F1					
	90% ; 10%	80% ; 20%	70% ; 30%	60% ; 40%	50% ; 50%	40% ; 60%
DTE-D3	0.838 / 0.801	0.806 / 0.767	0.798 / 0.761	0.799 / 0.762	0.803 / 0.768	0.798 / 0.762
DTE-D5	0.913 / 0.903	0.873 / 0.861	0.867 / 0.85	0.873 / 0.859	0.866 / 0.857	0.872 / 0.859
DTE-D8	0.960 / 0.961	0.957 / 0.957	0.956 / 0.956	0.941 / 0.940	0.925 / 0.925	0.926 / 0.924
DTE -D10	0.965 / 0.965	0.954 / 0.954	0.952 / 0.953	0.951 / 0.951	0.927 / 0.927	0.931 / 0.929
DTG-D3	0.838 / 0.801	0.806 / 0.767	0.798 / 0.761	0.799 / 0.762	0.803 / 0.768	0.798 / 0.762
DTG-D5	0.908 / 0.897	0.873 / 0.860	0.865 / 0.852	0.877 / 0.863	0.882 / 0.868	0.871 / 0.858

DTG-D8	0.965 / 0.967	0.939 / 0.939	0.917 / 0.917	0.915 / 0.914	0.921 / 0.920	0.927 / 0.925
DTG -D10	0.971 / 0.971	0.936 / 0.936	0.917 / 0.916	0.916 / 0.915	0.924 / 0.923	0.922 / 0.920

Gambar tree terbaik yang terbentuk DTG-D10



Dari hasil tersebut dapat disimpulkan bahwa semakin besar jumlah kedalaman maka hasil akurasi yang dihasilkan akan semakin besar atau semakin akurat sampai pada batasan tertentu ketika tidak ada lagi hasil yang belum jelas. Jika sudah sampai batasan tersebut, ketika kedalaman tree ditambah pun tidak akan meningkatkan akurasi dari Decision Tree. **Hasil maksimal diperoleh dengan menggunakan parameter data train sebanyak 90%, dengan depth 10, criterion : gini menghasilkan accuracy sebesar 0.971 dan f1 sebesar 0.971.**

SVM

Berikut merupakan hasil dari SVM dengan melakukan variasi pada kernel, C, dan Gamma yang digunakan, diambil 10 besar kombinasi terbaik

1. RBF

test_size	C	gamma	akurasi	f1
0.1	100	1.0	0.988	0.989
0.1	10	1.0	0.983	0.983
0.2	100	1.0	0.977	0.977
0.2	10	1.0	0.974	0.974
0.3	10	1.0	0.971	0.971
0.4	10	1.0	0.971	0.971
0.5	10	1.0	0.966	0.966
0.2	100	0.1	0.965	0.965
0.1	100	0.1	0.960	0.959
0.3	100	1.0	0.960	0.959

2. Poly

test_size	C	gamma	akurasi	f1
0.1	10	1.0	0.965	0.964
0.1	100	1.0	0.965	0.965
0.2	10	1.0	0.945	0.944
0.5	100	1.0	0.943	0.942
0.2	100	1.0	0.942	0.941
0.2	1	1.0	0.939	0.936
0.4	10	1.0	0.939	0.937
0.5	10	1.0	0.938	0.936
0.1	1	1.0	0.936	0.936
0.4	100	1.0	0.935	0.933

3. Sigmoid

test_size	C	gamma	akurasi	f1
0.1	100	0.01	0.884	0.880
0.3	100	0.1	0.882	0.880
0.5	100	0.01	0.882	0.877
0.3	100	0.01	0.881	0.876
0.4	100	0.01	0.881	0.877
0.5	10	0.1	0.881	0.876
0.1	10	0.1	0.879	0.875
0.3	10	0.1	0.879	0.874

0.4	10	0.1	0.877	0.873
0.2	10	0.1	0.876	0.871

4. Linear

test_size	C	gamma	akurasi	f1
0.2	10	0.1	0.902	0.899
0.2	10	0.01	0.902	0.899
0.2	10	0.001	0.902	0.899
0.2	10	1.0	0.902	0.899
0.3	100	0.1	0.900	0.897
0.3	100	0.01	0.900	0.897
0.3	100	0.001	0.900	0.897
0.3	100	1.0	0.900	0.897
0.2	100	0.1	0.899	0.896
0.2	100	0.01	0.899	0.896

Diperoleh hasil data terbaik dengan menggunakan kombinasi 90% data train, C: 100, gamma: 1.0, kernel: rbf dengan accuracy sebesar 0.988 dan f1 sebesar 0.988. Jadi dapat kita lihat bahwa penggunaan data train terbanyak yakni 90% dan 10% data test dan dengan menggunakan kernel RBF akan menghasilkan akurasi terbaik. Dalam kernel linear akurasi terbaik didapatkan dari data train sebanyak 80% dan data test sebanyak 20%. Hal ini bisa terjadi karena kurangnya pemilihan data yang digunakan sebagai data train, outlier tidak dihilangkan, dan algoritma dari kernel linear masih kurang kompleks.