Hasil *Statistical Learning* untuk pembahasan.

Muhammad Kevin A (19/445592/PA/19416)

Referensi:

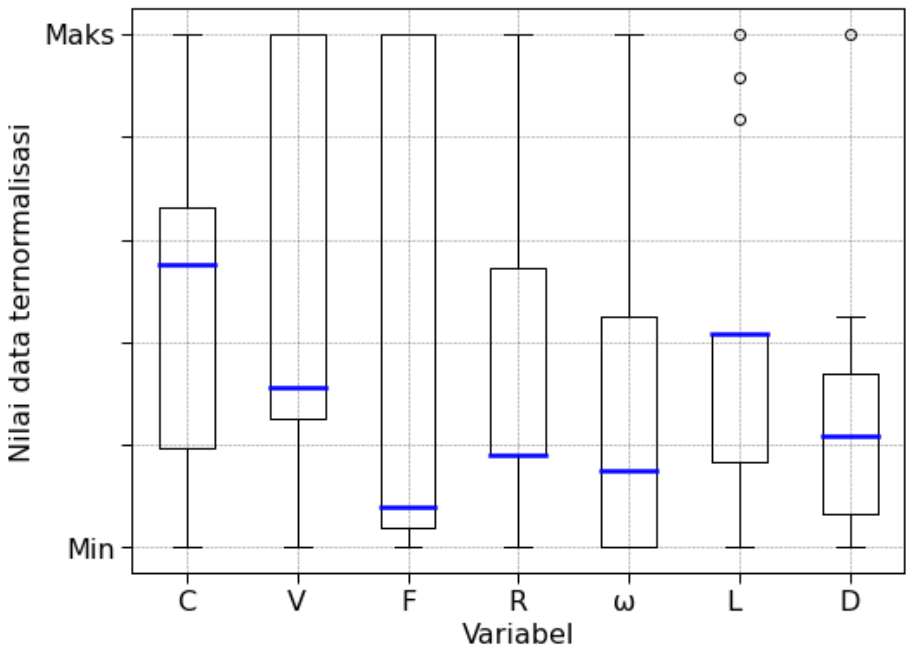
- The Elements of Statistical Learning - Google Books
- Statistical Learning with Math and Python - Google Books
- HSI2020_takigawa1_highres (itakigawa.github.io)

A. EKSPLORASI DATA

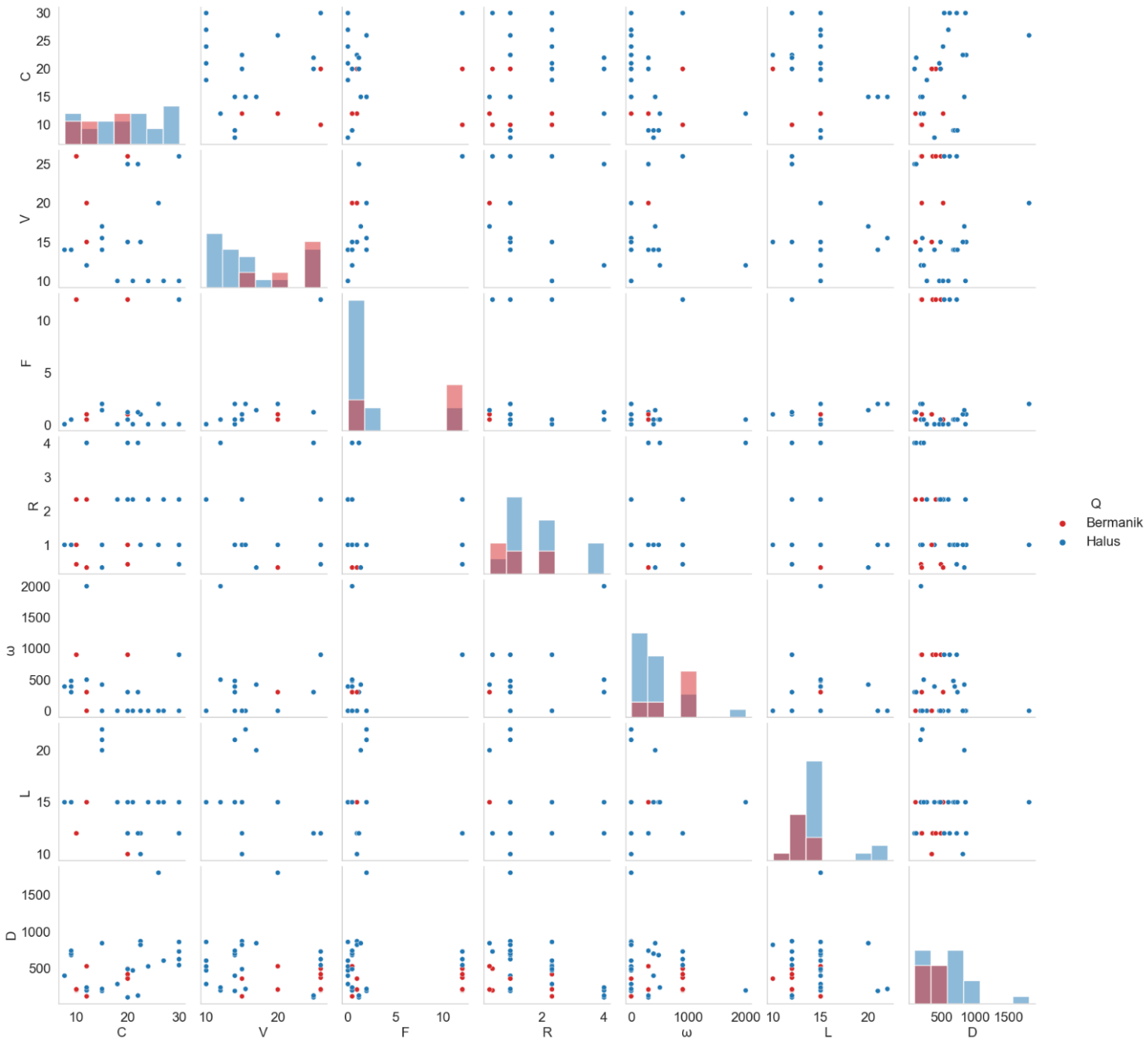
1. Statistika Deskriptif

	Jumlah	Rata-rata	Simpangan Baku	Min	Kuartil 1	Median	Kuartil 3	Maks
C	33.0	17.96	7.06	7.70	12.0	20.00	22.50	30.0
V	33.0	18.05	6.15	10.00	14.0	15.00	26.00	26.0
F	33.0	3.82	5.11	0.05	0.5	1.00	12.00	12.0
R	33.0	1.66	1.14	0.33	1.0	1.00	2.33	4.0
ω	33.0	417.58	465.93	0.00	0.0	300.00	900.00	2000.0
L	33.0	14.15	2.76	10.00	12.0	15.00	15.00	22.0
D	33.0	491.31	334.91	105.00	215.0	471.58	680.00	1800.0

2. Visualisasi Sebaran Data



3. Visualisasi Hubungan Antar Variabel



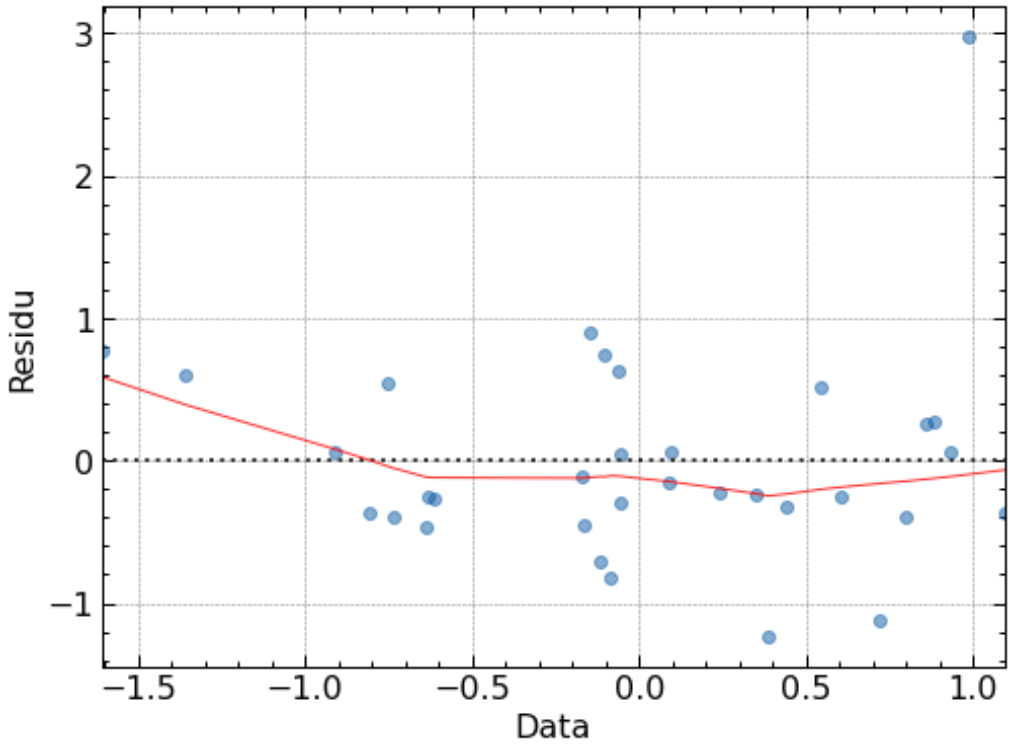
B. REGRESI LINEAR

1. Regresi Linear Sederhana

Variabel Keluaran:	D
Jumlah Variabel Masukan:	6
Model:	Regresi linear sederhana
Metode:	RSS
Jumlah Observasi:	33
R^2:	0.461

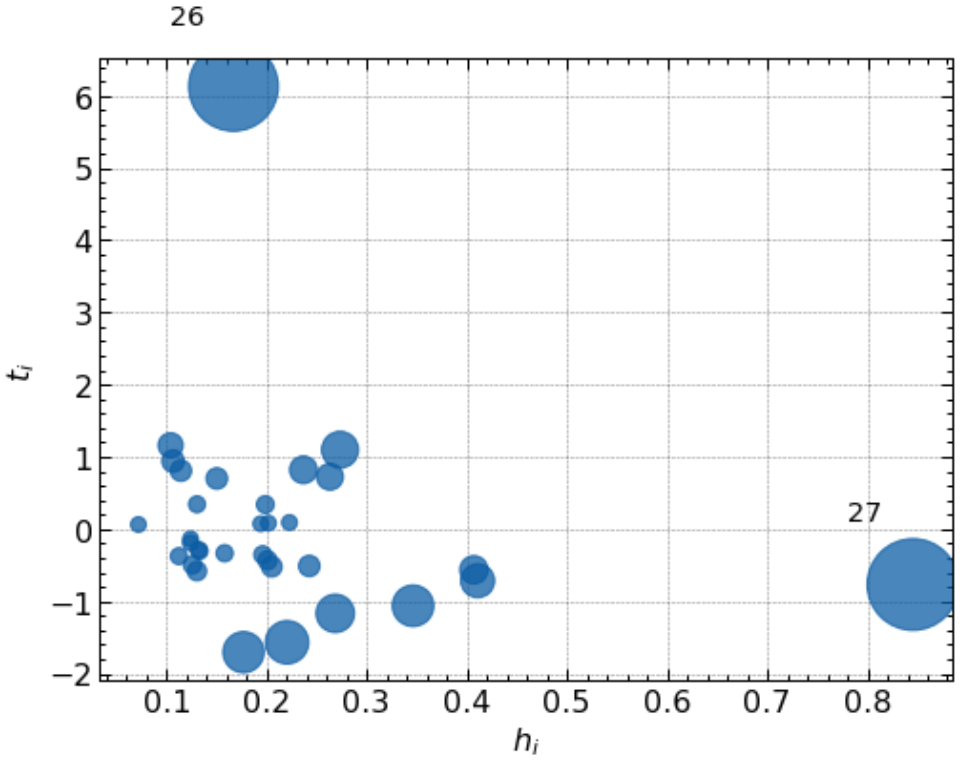
	Koefisien	σ	t-statistik	P>|t|	Interval Konfidensi	
					[0.025	0.975]
Intersep	1.804e-16	0.144	1.25e-15	1.000	-0.296	0.296
C	0.5931	0.170	3.492	0.002	0.244	0.942
V	-0.0178	0.272	-0.065	0.948	-0.577	0.541
F	-0.5163	0.328	-1.572	0.128	-1.191	0.159
R	-0.5895	0.169	-3.480	0.002	-0.938	-0.241
ω	0.2222	0.219	1.012	0.321	-0.229	0.673
L	-0.0679	0.174	-0.390	0.700	-0.426	0.290

2. Plot Residu



3. Deteksi *outlier* menggunakan nilai h_i, t_i, & D_i

h_i : Titik leverage (Sumbu horizontal)
t_i : Titik *studentized residual* (Sumbu vertikal)
D_i: Jarak Cook (Semakin besar lingkaran, semakin besar nilai **D_i**)



Data Outlier:

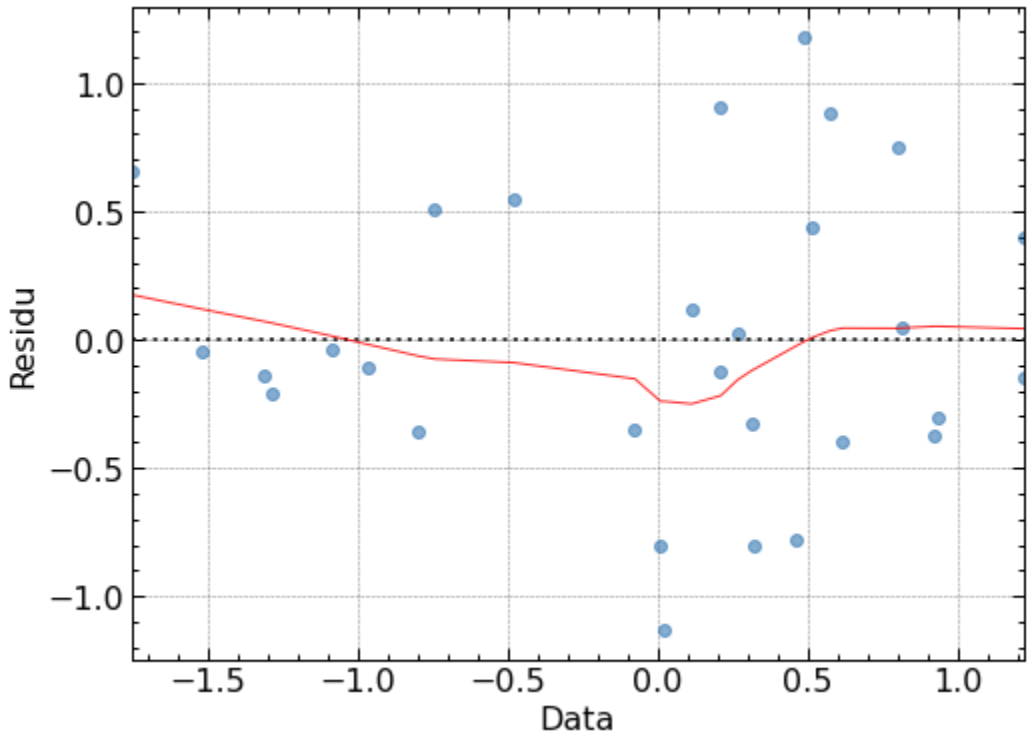
C	V	F	R	ω	L	D	Q	Sumber PET	Referensi
26	20	2	1	0	15	1800	Halus	Botol bekas PET (Sharalau)	Khorram *et al.*, 2017
12	12	0.5	4	2000	15	200	Halus	Butiran PET (IPRT)	Abassi *et al.*, 2018

4. Regresi linear sederhana setelah data *outlier* dihapus.

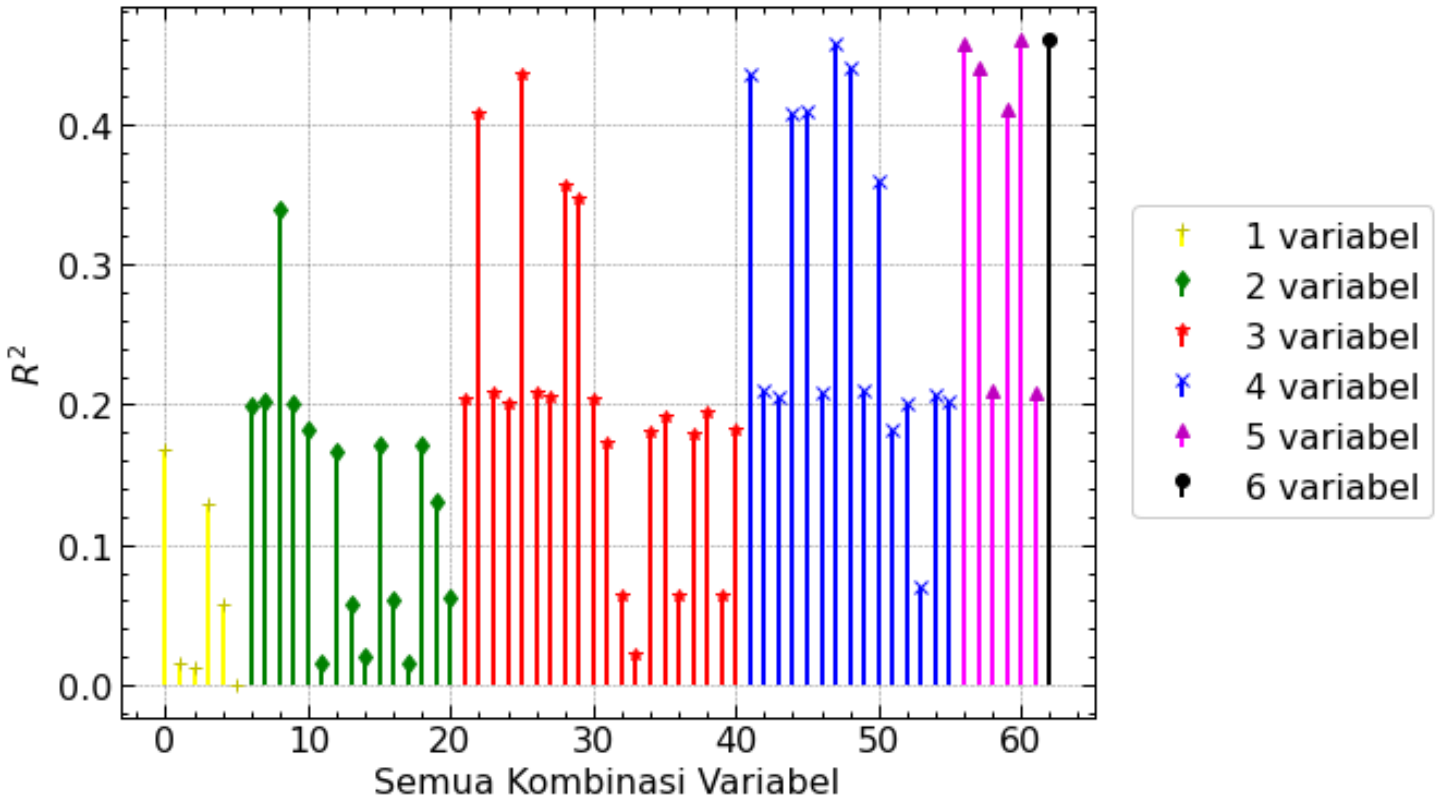
Variabel Keluaran:	D
Jumlah Variabel Masukan:	6
Model:	Regresi linear sederhana
Metode:	RSS
Jumlah Observasi:	31
R^2:	0.693

	Koefisien	σ	t-statistik	P>|t|	Interval Konfidensi	
					[0.025	**0.975]**
const	2.047e-16	0.113	1.81e-15	1.000	-0.234	0.234
C	0.7211	0.142	5.078	0.000	0.428	1.014
V	-0.6509	0.229	-2.847	0.009	-1.123	-0.179
F	-0.9328	0.327	-2.851	0.009	-1.608	-0.258
R	-0.6145	0.125	-4.914	0.000	-0.873	-0.356
ω	1.1324	0.308	3.682	0.001	0.498	1.767
L	-0.1720	0.138	-1.248	0.224	-0.457	0.113

Plot residu setelah data *outlier* dihapus

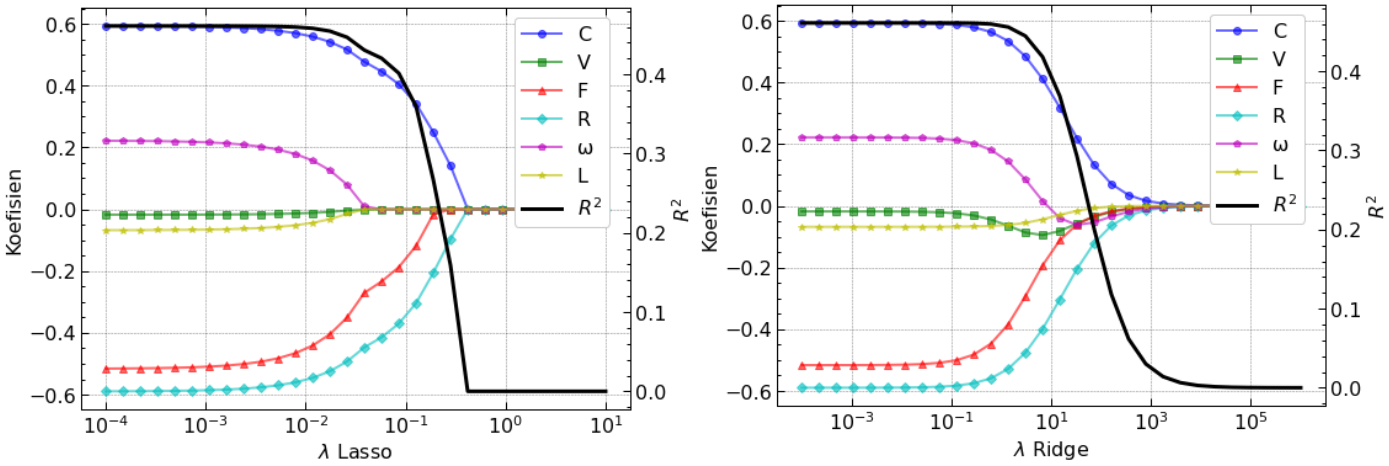


5. Regresi linear sederhana untuk semua kombinasi variabel yang mungkin



Kombinasi Variabel	R^2
(C, V, F, R, ω, L)	0.461135
(C, F, R, ω, L)	0.461046
(C, V, F, R, ω)	0.457986
(C, F, R, ω)	0.457969
(C, V, F, R, L)	0.439904
(C, F, R, L)	0.439478
(C, V, F, R)	0.436097
(C, F, R)	0.436058
(C, V, R, ω, L)	0.40993
(C, V, R, L)	0.409392

6. Regresi linear sederhana dengan regularisasi



Tabel Koefisien Regresi Lasso

Lambda	**0.010**	**0.025**	**0.040**	**0.055**	**0.070**	**0.085**	**0.100**
C	0.565	0.520	0.472	0.450	0.427	0.405	0.382
V	0.000	0.000	0.000	0.000	0.000	0.000	0.000
F	-0.463	-0.360	-0.263	-0.238	-0.213	-0.188	-0.163
R	-0.552	-0.496	-0.442	-0.418	-0.394	-0.371	-0.347
ω	0.168	0.085	0.001	0.000	0.000	0.000	0.000

L	-0.046	-0.016	0.000	0.000	0.000	0.000	0.000
R^2	**0.459**	**0.448**	**0.429**	**0.422**	**0.413**	**0.402**	**0.389**

7. Regularisasi regresi linear dengan suku interaksi

Variabel Keluaran:	D
Jumlah Variabel Masukan:	21
Model:	Regresi linear dengan suku interaksi
Metode:	RSS
Jumlah Observasi:	33
R^2:	0.990

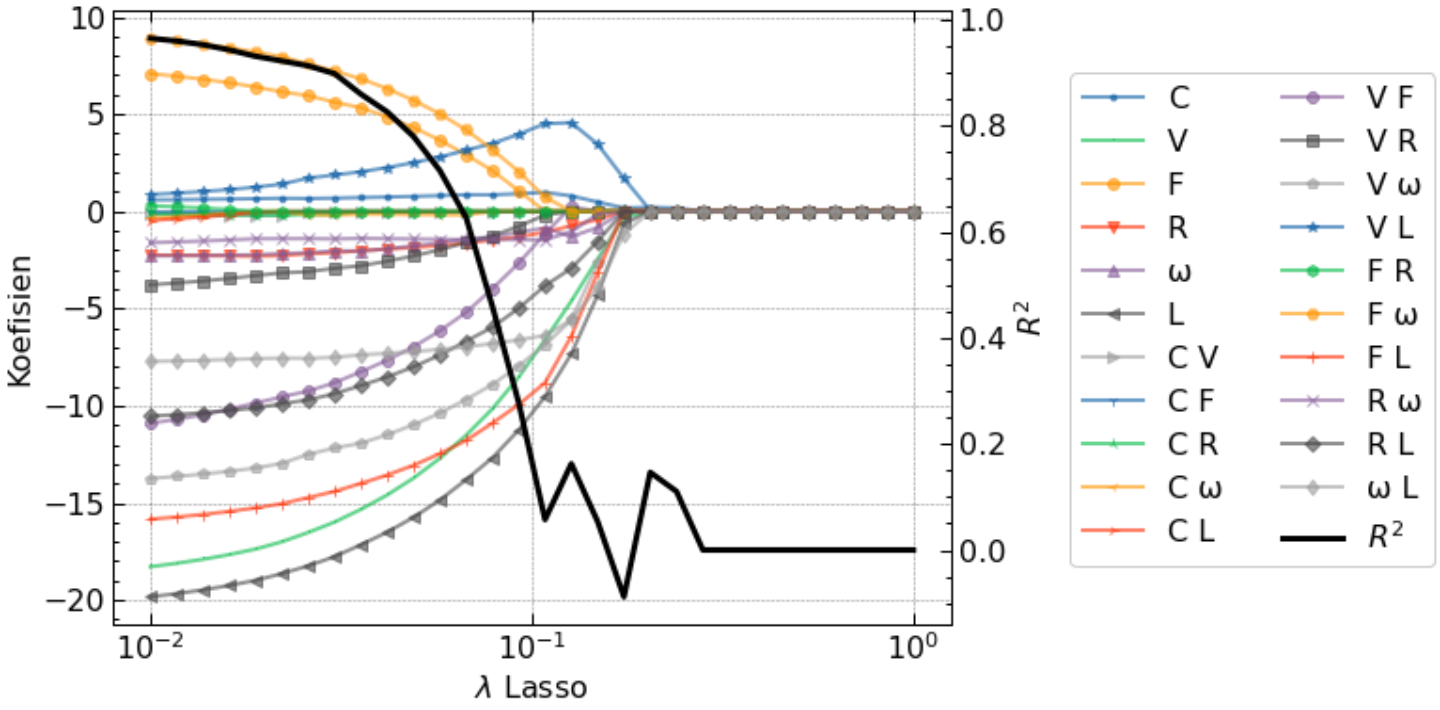
Catatan penting:

Problems with R-squared statistic : Its value never decreases no matter the number of variables we add to our regression model. That is, even if we are adding redundant variables to the data, the value of R-squared does not decrease. It either remains the same or increases with the addition of new independent variables.

Coefficient of determination - Wikipedia

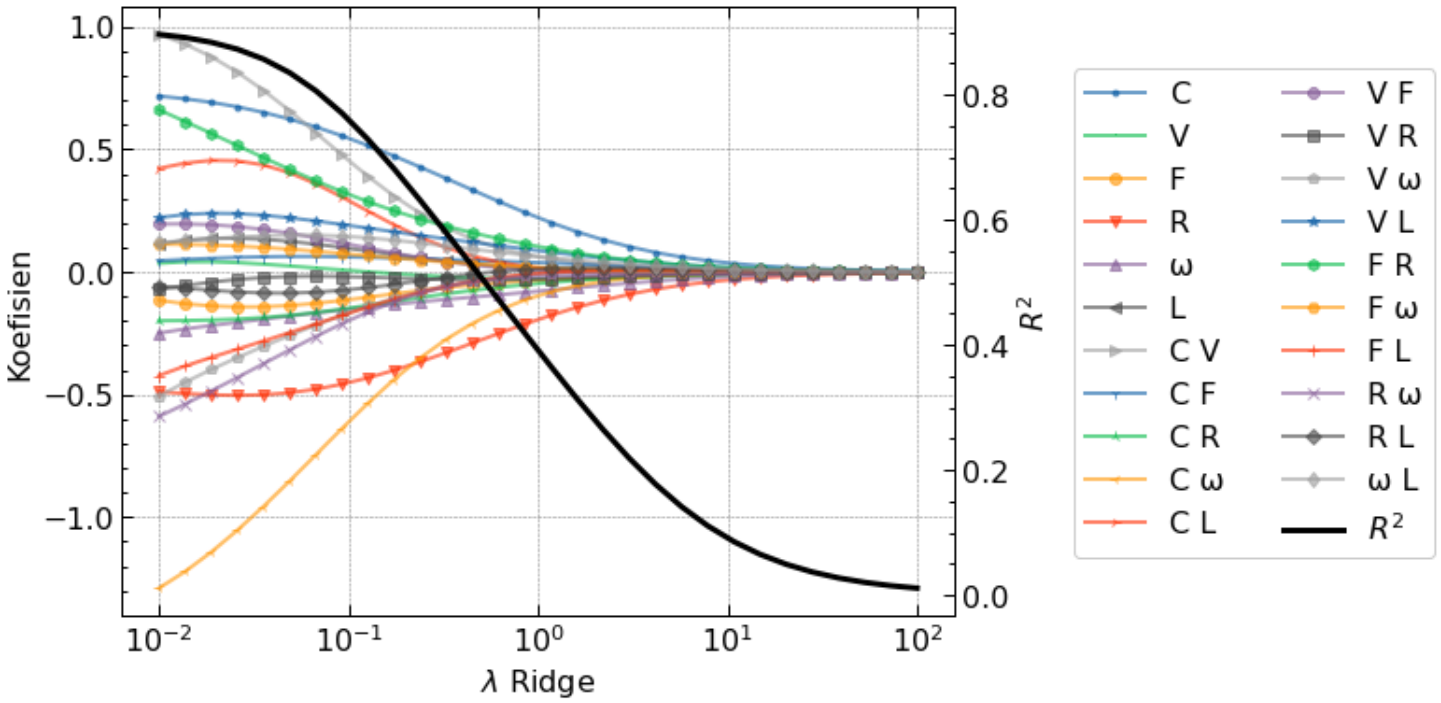
	Koefisien	σ	t-statistik	P>|t|	Interval Konfidensi	
					[0.025	**0.975]**
const	0	0.031	0	1.000	-0.067	0.067
C	8.0018	2.219	3.606	0.004	3.118	12.886
V	40.0545	10.523	3.806	0.003	16.893	63.216
F	29.8049	32.793	0.909	0.383	-42.371	101.981
R	113.6203	15.980	7.110	0.000	78.449	148.792
ω	126.3254	17.036	7.415	0.000	88.829	163.822
L	30.0297	4.635	6.478	0.000	19.827	40.232
C V	-0.8805	1.241	-0.710	0.493	-3.611	1.850
C F	0.6480	3.065	0.211	0.836	-6.097	7.393
C R	-0.3332	0.213	-1.565	0.146	-0.802	0.135
C ω	-2.6290	2.966	-0.886	0.394	-9.158	3.900

C L	-5.8562	1.784	-3.282	0.007	-9.783	-1.929
V F	-54.0295	12.693	-4.257	0.001	-81.967	-26.092
V R	-28.1357	4.400	-6.394	0.000	-37.821	-18.451
V ω	-103.2591	17.249	-5.986	0.000	-141.225	-65.294
V L	-15.3784	5.751	-2.674	0.022	-28.037	-2.720
F R	3.3562	0.999	3.359	0.006	1.157	5.555
F ω	70.5901	16.865	4.186	0.002	33.471	107.710
F L	-19.0289	19.995	-0.952	0.362	-63.037	24.979
R ω	-9.7382	1.734	-5.615	0.000	-13.555	-5.921
R L	-89.7022	12.412	-7.227	0.000	-117.022	-62.383
ω L	-52.6762	6.652	-7.919	0.000	-67.316	-38.036

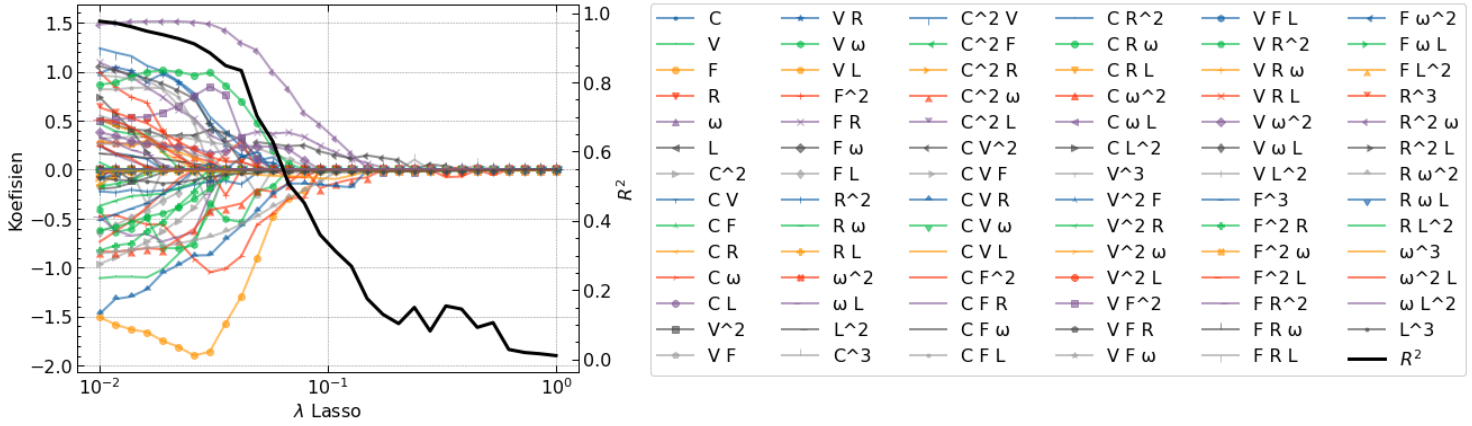


Lambda	**0.010**	**0.025**	**0.040**	**0.055**	**0.070**	**0.085**	**0.100**
C	0.586	0.652	0.726	0.796	0.866	0.875	0.940
V	-18.282	-16.631	-14.805	-12.996	-11.186	-9.414	-7.606
F	7.081	6.025	4.996	3.845	2.697	1.618	0.470

R	-2.251	-2.195	-1.970	-1.788	-1.607	-1.402	-1.217
ω	-2.290	-2.152	-1.960	-1.661	-1.365	-1.208	-0.919
L	-19.830	-18.333	-16.691	-15.104	-13.516	-12.029	-10.443
C V	-0.130	0.000	0.000	0.000	0.000	0.000	0.000
C F	0.000	0.000	0.000	0.000	0.000	0.000	0.000
C R	-0.150	-0.152	0.000	0.000	0.000	0.000	0.000
C ω	-0.358	-0.183	-0.156	-0.177	-0.199	0.000	0.000
C L	-0.471	0.000	0.000	0.000	0.000	0.000	0.000
V F	-10.904	-9.336	-7.833	-6.383	-4.933	-3.382	-1.926
V R	-3.779	-3.143	-2.635	-2.068	-1.501	-1.043	-0.482
V ω	-13.741	-12.570	-11.618	-10.550	-9.484	-8.474	-7.411
V L	0.861	1.672	2.187	2.713	3.237	3.705	4.226
F R	0.288	0.000	0.000	0.000	0.000	0.000	0.000
F ω	8.899	7.662	6.453	5.208	3.962	2.654	1.408
F L	-15.839	-14.787	-13.690	-12.640	-11.590	-10.461	-9.408
R ω	-1.615	-1.404	-1.403	-1.436	-1.469	-1.443	-1.471
R L	-10.538	-9.766	-8.648	-7.566	-6.483	-5.518	-4.438
ω L	-7.725	-7.576	-7.329	-7.136	-6.939	-6.717	-6.516
R2	0.965	0.916	0.837	0.735	0.600	0.380	0.163



8. Regularisasi Lasso untuk regresi linear dengan suku polinomial derajat 3



Regresi Lasso untuk lambda = 0.04

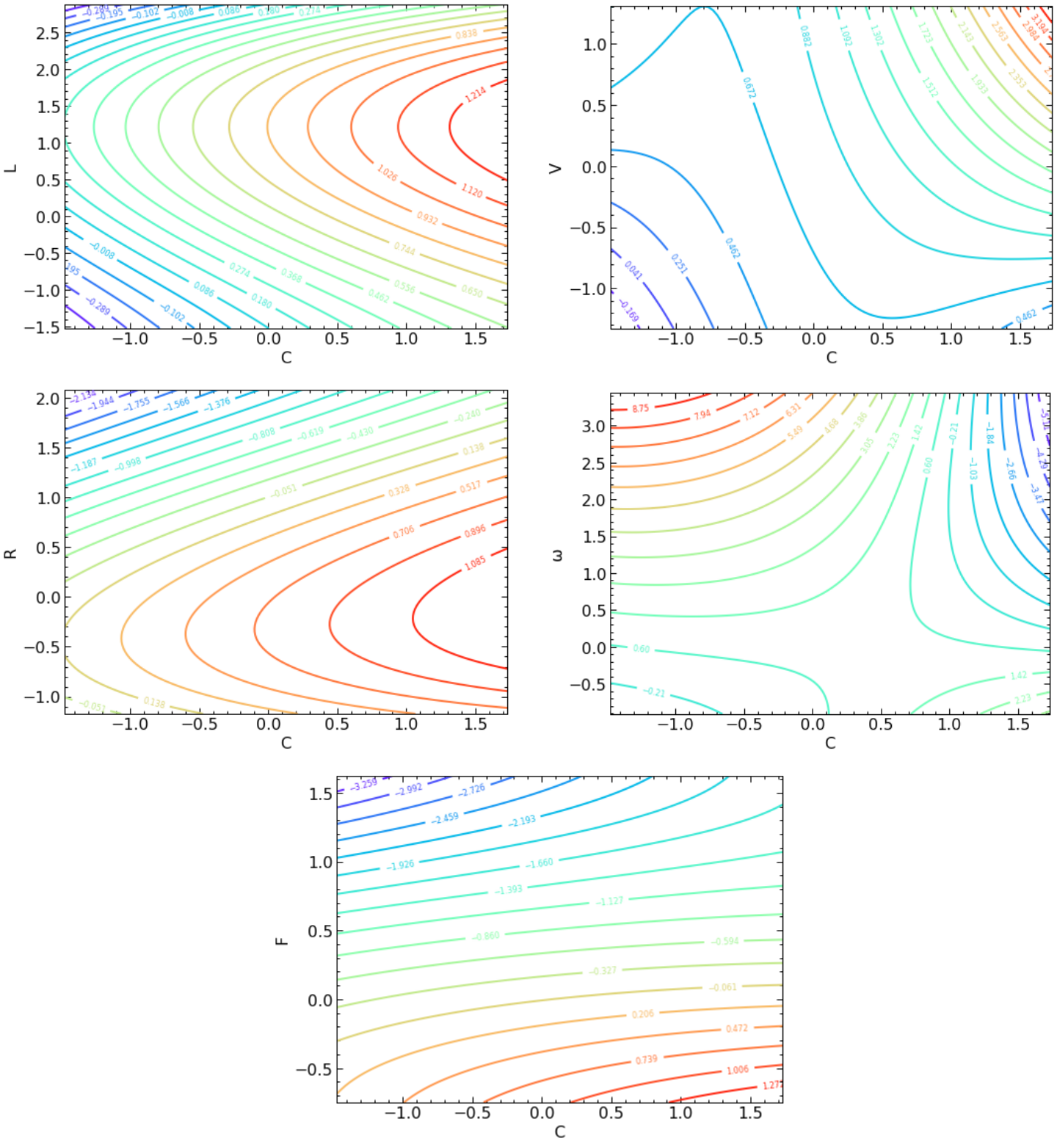
Variabel Keluaran:	D
Jumlah Variabel Masukan:	30
Model:	Regresi linear dengan suku polinomial orde 3
Metode:	RSS
Jumlah Observasi:	30
R^2:	0.840

	coef	std err	t	P>|t|	[0.025	0.975]
F	-1.3030	3.214	-0.405	0.724	-15.133	12.527
L	0.3292	1.560	0.211	0.852	-6.382	7.041
C ω	-0.8580	1.850	-0.464	0.688	-8.816	7.100
V R	0.1888	1.557	0.121	0.915	-6.509	6.886
V ω	-0.4568	1.770	-0.258	0.820	-8.073	7.159
F R	0.3559	4.671	0.076	0.946	-19.744	20.456
R^2	-0.1452	1.278	-0.114	0.920	-5.645	5.354
L^2	-0.0097	0.953	-0.010	0.993	-4.110	4.090
C^2 V	0.2732	11.306	0.024	0.983	-48.373	48.919
C^2 ω	-0.4010	6.509	-0.062	0.956	-28.408	27.607
C V^2	0.0549	8.107	0.007	0.995	-34.825	34.935
C V F	0.2958	3.849	0.077	0.946	-16.263	16.855
C V R	-0.5890	8.452	-0.070	0.951	-36.953	35.775
C F^2	0.2395	6.725	0.036	0.975	-28.696	29.175
C R ω	0.7383	3.399	0.217	0.848	-13.887	15.363
V F^2	0.3584	4.970	0.072	0.949	-21.027	21.744
V ω^2	0.1972	6.719	0.029	0.979	-28.713	29.107
V L^2	0.1119	5.118	0.022	0.985	-21.910	22.133
F^3	0.0885	2.204	0.040	0.972	-9.396	9.573
R^3	0.0863	3.847	0.022	0.984	-16.464	16.636
R^2 ω	1.3335	2.827	0.472	0.684	-10.830	13.497
R ω^2	-0.6226	15.377	-0.040	0.971	-66.785	65.540
L^3	-0.0381	0.768	-0.050	0.965	-3.342	3.266

Model regresi polinomial orde 3

-1.303*F +0.3292*L +-0.858*C ω +0.1888*V R +-0.4568*V ω +0.3559*F R +-0.1452*R^2 +-0.0097*L^2 +0.2732*C^2 V +-0.401*C^2 ω +0.0549*C V^2 +0.2958*C V F +-0.589*C V R +0.2395*C F^2 +0.7383*C R ω +0.3584*V F^2 +0.1972*V ω^2 +0.1119*V L^2 +0.0885*F^3 +0.0863*R^3 +1.3335*R^2 ω +-0.6226*R ω^2 +-0.0381*L^3

Plot kontur untuk model regresi polinomial orde 3 (satuan dalam simpangan baku)



C. REGRESI LOGISTIK

	coef	std err	z	P>|z|	[0.025	0.975]
Intercept	3.3576	1.676	2.004	0.045	0.073	6.642
C	3.8344	2.003	1.914	0.056	-0.092	7.760
V	-2.8344	1.759	-1.611	0.107	-6.283	0.614
F	-0.2089	1.341	-0.156	0.876	-2.837	2.419
R	1.1778	0.916	1.286	0.198	-0.617	2.972
ω	0	0	nan	nan	0	0
L	3.5695	2.132	1.674	0.094	-0.610	7.74

	C	V	F	R	ω	L	Halus	prob_halus(%)
0	-1.145211	1.313551	1.624006	-1.088160	1.051462	-0.790844	0	0.010
1	-1.145211	1.313551	1.624006	-0.581206	1.051462	-0.790844	0	0.018
2	-1.145211	1.313551	1.624006	0.601689	1.051462	-0.790844	0	0.074
3	0.293387	1.313551	1.624006	-1.088160	1.051462	-0.790844	0	2.450
4	0.293387	1.313551	1.624006	-0.581206	1.051462	-0.790844	0	4.364
5	0.293387	1.313551	1.624006	0.601689	1.051462	-0.790844	0	15.525
6	1.731985	1.313551	1.624006	-1.088160	1.051462	-0.790844	1	86.197
7	1.731985	1.313551	1.624006	-0.581206	1.051462	-0.790844	1	91.900
8	1.731985	1.313551	1.624006	0.601689	1.051462	-0.790844	1	97.859
9	-1.476089	-0.668035	-0.748938	-0.581206	-0.060102	0.311882	1	54.396
10	-1.289071	-0.668035	-0.659580	-0.581206	0.136056	0.311882	1	70.573
11	-1.289071	-0.668035	-0.659580	-0.581206	-0.060102	0.311882	1	70.573
12	-1.289071	-0.668035	-0.659580	-0.581206	-0.256261	0.311882	1	70.573
13	0.653036	-0.502902	-0.560294	-0.581206	-0.910122	-0.790844	1	98.009
14	0.005667	-1.328563	-0.748422	0.601689	-0.910122	0.311882	1	99.989
15	0.437247	-1.328563	-0.748422	0.601689	-0.910122	0.311882	1	99.998

16	0.868826	-1.328563	-0.748422	0.601689	-0.910122	0.311882	1	100.000
17	1.300405	-1.328563	-0.748422	0.601689	-0.910122	0.311882	1	100.000
18	1.731985	-1.328563	-0.748422	0.601689	-0.910122	0.311882	1	100.000
19	0.293387	-0.502902	-0.560294	-0.581206	-0.910122	-1.525995	0	47.336
20	0.653036	-0.502902	-0.560294	-0.581206	-0.910122	-1.525995	1	78.115
21	0.293387	1.148419	-0.520580	2.080306	-0.256261	-0.790844	1	72.382
22	0.581106	1.148419	-0.520580	2.080306	-0.256261	-0.790844	1	88.762
23	-0.857492	0.322758	-0.659580	-1.172653	-0.256261	0.311882	0	27.382
24	-0.857492	0.322758	-0.560294	-1.172653	-0.256261	0.311882	0	26.972
25	-0.425912	-0.172638	-0.480865	-1.172653	0.005284	2.149759	1	99.982
26	1.156546	0.322758	-0.361721	-0.581206	-0.910122	0.311882	1	99.938
27	-0.857492	-0.998299	-0.659580	2.080306	3.448953	0.311882	1	99.864
28	-0.857492	-0.998299	-0.659580	2.080306	0.179647	0.311882	1	99.864
29	-0.425912	-0.668035	-0.361721	-0.581206	-0.910122	2.517334	1	99.999
30	-0.425912	-0.420336	-0.361721	-0.581206	-0.910122	2.884910	1	100.000
31*	-0.857492	-0.502902	-0.659580	0.601689	-0.910122	0.311882	0	96.937
32	0.293387	-0.502902	-0.659580	0.601689	-0.910122	0.311882	1	99.962

***Data Anomali**

Plot probabilitas untuk 3 variabel yang paling signifikan