

TUGAS
Evaluation Method



Anggota Kelompok:

Gerald Nathanael - 11S21015
Boy Martahan Sitorus - 11S21025
Dastin Raynold Sirait - 11S21037
Gerry Benyamin Abdiel Bukit - 11S21055

Institut Teknologi Del
Sitoluama, Laguboti, Toba Samosir
Sumatera Utara, Indonesia

Pembagian Tugas:

Gerald Nathanael	FOLD 3 - K = 5 EUCLIDEAN
Boy Martahan Sitorus	FOLD 1 - K = 5 EUCLIDEAN
Dastin Raynold Sirait	FOLD 4- K = 5 EUCLIDEAN
Gerry Benyamin Abdiel Bukit	FOLD 2- K = 5 EUCLIDEAN

Sumarry Laporan:

Detail Perhitungan:

A. Standarisasi Data

Pada tahap pertama kita perlu melakukan standarisasi pada dataset kita, kita dapat melakukan standarisasi dengan menggunakan formula :

$$Xs = X - \text{mean} / \text{Max} - \text{Min}$$

Sehingga akan menghasilkan table sebagai berikut :

Standardized Dataset						
ID	studytime	freetime	goout	absences	graduation	
1	-0.025	-0.0625	0.325	0.125	Fail	FOLD 1 (Boy Sitorus)
2	0.475	-0.3125	-0.175	0	Pass	
3	-0.025	0.1875	-0.425	-0.25	Pass	
4	-0.025	-0.0625	0.075	-0.25	Fail	
5	-0.025	-0.0625	0.075	0	Fail	
6	-0.025	0.4375	-0.425	-0.25	Pass	FOLD 2 (Gerry Bukit)
7	0.475	-0.3125	0.075	0.125	Pass	
8	-0.025	0.1875	0.075	-0.125	Fail	
9	-0.025	-0.0625	-0.175	0.375	Fail	
10	-0.025	-0.3125	-0.175	-0.25	Pass	FOLD 3 (Gerald Nathanael)
11	-0.525	0.1875	0.325	0	Pass	
12	-0.025	-0.0625	-0.175	0	Fail	
13	-0.025	-0.0625	-0.175	0	Fail	
14	-0.025	0.1875	-0.175	0.375	Pass	FOLD 4 (Dastin Sirait)
15	0.475	0.4375	-0.175	-0.25	Pass	
16	-0.525	0.4375	0.575	0.75	Fail	
17	-0.025	0.1875	0.325	-0.25	Fail	
18	0.475	-0.3125	-0.175	-0.125	Pass	
19	-0.525	-0.0625	0.075	-0.125	Pass	
20	-0.025	-0.5625	0.325	0.125	Fail	

B. Fold - 1 KKN(K = 5) Euclidean Distance

Untuk Fold - 1, kita akan mencari euclidean distance antara ID 1, 2, 3, 4 dengan ID lainnya

FOLD 1 - K = 5, EUCLIDEAN DISTANCE KNN					
Distance To ID	Graduation	1	2	3	4
5	Fail	0.2795084972	0.6123724357	0.6123724357	0.25
6	Pass	0.9762812095	0.9682458366	0.25	0.7071067812
7	Pass	0.6123724357	0.2795084972	0.9437293044	0.6731456009
8	Fail	0.4330127019	0.7603453163	0.5153882032	0.2795084972
9	Fail	0.5590169944	0.6731456009	0.7180703308	0.6731456009
10	Pass	0.6731456009	0.5590169944	0.5590169944	0.3535533906
11	Pass	0.5728219619	1.224744871	0.9354143467	0.6614378278
12	Fail	0.5153882032	0.5590169944	0.4330127019	0.3535533906
13	Fail	0.5153882032	0.5590169944	0.4330127019	0.3535533906
14	Pass	0.6123724357	0.8003905297	0.6731456009	0.7180703308
15	Pass	0.9437293044	0.790569415	0.6123724357	0.75
16	Fail	0.9762812095	1.639359631	1.520690633	1.322875656

Lalu kita akan melakukan validasi pada ID 1, 2, 3, 4 dengan menggunakan K = 5 pada KKN, maka akan dihasilkan table confusion matrix sebagai berikut :

ID	Target	Actual		Target	Prediction	
1	Fail	Fail			Pass	Fail
2	Pass	Fail		Pass	0	2
3	Pass	Fail		Fail	0	2
4	Fail	Fail		Classification Acc	0.5	

C. Fold - 2 KKN(K = 5) Euclidean Distance

Untuk Fold - 2, kita akan mencari euclidean distance antara ID 5, 6, 7, 8 dengan ID lainnya

FOLD 2- K = 5, EUCLIDEAN DISTANCE KNN					
Distance To ID	Graduation	5	6	7	8
1	Fail	0.2795084972	0.9960076556	0.5722761571	0.2795084972
2	Pass	0.6123724357	0.9975751851	0.990265116	0.58630197
3	Pass	0.6123724357	0.2875	0.9182183836	0.7288689869
4	Fail	0.25	0.734102343	0.6368869601	0.3952847075
9	Fail	0.5590169944	0.8614123577	0.6244997998	0.375
10	Pass	0.6731456009	0.8262301435	0.6368869601	0.4677071733
11	Pass	0.5728219619	0.9785991263	0.575	0.6846531969
12	Fail	0.5153882032	0.6433554616	0.5857687257	0.3061862178
13	Fail	0.5153882032	0.6433554616	0.5857687257	0.3061862178
14	Pass	0.6123724357	0.7319707986	0.7599342077	0.5153882032
15	Pass	0.9437293044	0.5602733708	1.291559135	0.9185586535
16	Fail	0.9762812095	1.500468677	1.098009563	1.131923142

Lalu kita akan melakukan validasi pada ID 5, 6, 7, 8 dengan menggunakan K = 5 pada KKN, maka akan dihasilkan table confusion matrix sebagai berikut :

ID	Target	Actual		Target	Prediction	
5	Fail	Fail			Pass	Fail
6	Pass	Pass		Pass	1	1
7	Pass	Fail		Fail	0	2
8	Fail	Fail		Classification Acc	0.75	

D. Fold - 3 KKN(K = 5) Euclidean Distance

Untuk Fold - 3, kita akan mencari euclidean distance antara ID 9, 10, 11, 12 dengan ID lainnya

FOLD 3- K = 5, EUCLIDEAN DISTANCE KNN					
Distance To ID	Graduation	9	10	11	12
1	Fail	0.5590169944	0.6731456009	0.5303300859	0.5153882032
2	Pass	0.6731456009	0.5590169944	1.125	0.5590169944
3	Pass	0.7180703308	0.5590169944	1.007782219	0.4330127019
4	Fail	0.6731456009	0.3535533906	0.625	0.3535533906
5	Fail	0.4506939094	0.4330127019	0.5728219619	0.25
6	Pass	0.8385254916	0.790569415	1.125	0.6123724357
7	Pass	0.6614378278	0.6731456009	1.045825033	0.625
8	Fail	0.6123724357	0.5728219619	0.6846531969	0.375
13	Fail	0.375	0.3535533906	0.7180703308	0
14	Pass	0.25	0.8003905297	0.8838834765	0.4506939094
15	Pass	0.9437293044	0.9013878189	1.305038314	0.75
16	Fail	1.096870548	1.541103501	1.007782219	1.274754878

Lalu kita akan melakukan validasi pada ID 9, 10, 11, 12 dengan menggunakan K = 5 pada KKN, maka akan dihasilkan table confusion matrix sebagai berikut :

ID	Target	Actual	Target	Prediction	
9	Fail	Fail		Pass	Fail
10	Pass	Fail	Pass	0	2
11	Pass	Fail	Fail	0	2
12	Fail	Fail	Classification Acc	0.5	

E. Fold - 4 KKN(K = 5) Euclidean Distance

Untuk Fold - 4, kita akan mencari euclidean distance antara ID 13, 14, 15, 16 dengan ID lainnya

FOLD 4- K = 5, EUCLIDEAN DISTANCE KNN					
Distance To ID	Graduation	13	14	15	16
1	Fail	0.5153882032	0.6123724357	0.9437293044	0.9762812095
2	Pass	0.5590169944	0.8003905297	0.790569415	1.639359631
3	Pass	0.4330127019	0.6731456009	0.6123724357	1.520690633
4	Fail	0.3535533906	0.7180703308	0.75	1.322875656
5	Fail	0.25	0.5153882032	0.790569415	1.145643924
6	Pass	0.6123724357	0.7180703308	0.5590169944	1.5
7	Pass	0.625	0.790569415	0.875	1.484292761
8	Fail	0.375	0.5590169944	0.625	1.152443057
9	Fail	0.375	0.25	0.9437293044	1.096870548
10	Pass	0.3535533906	0.8003905297	0.9013878189	1.541103501
11	Pass	0.75	0.8003905297	1.17260394	0.8291561976
12	Fail	0	0.4506939094	0.75	1.274754878

Lalu kita akan melakukan validasi pada ID 13, 14, 15, 16 dengan menggunakan K = 5 pada KKN, maka akan dihasilkan table confusion matrix sebagai berikut :

ID	Target	Actual	Target	Prediction	
13	Fail	Fail		Pass	Fail
14	Pass	Fail	Pass	0	2
15	Pass	Fail	Fail	0	2
16	Fail	Fail	Classification Acc	0.5	

Kesimpulan:

Setelah dilakukan standarisasi, dan melakukan K-FOLD dengan $K = 4$, dan KNN($K=5$) dengan Euclidean Distance, kita dapat memperoleh bahwa Fold dengan accuracy tertinggi yang dihasilkan oleh KNN($K=5$) adalah Fold - 4.

Link Spredsheets:

https://docs.google.com/spreadsheets/d/1IARkx9zX7bjzaex_nQPHKaXh9lcEuAjpAaJh93WHd5c/edit#gid=1439030261