

BAB V

ANALISA DAN PEMBAHASAN

5.1 Pengujian

Pada subbab ini akan dilakukan pembahasan mengenai pengujian yang telah dilakukan pada sistem dan hasil evaluasi dari sistem tersebut. Didalam pengujian ini digunakan beberapa *dataset* yang semua atributnya berdominan dengan data kategori. *Dataset* yang digunakan sebagai ujicoba adalah *dataset* yang diambil dari *UCI Machine Learning Repository*. Berikut adalah *dataset* yang di pakai dalam pengujian.

Tabel 5.1 *Dataset* yang digunakan

No.	Nama Dataset	Jumlah Data	Jumlah Atribut	Jumlah Class
1.	<i>Balance-Scale</i>	625	4	3
2.	<i>Car Evaluation</i>	1728	6	4
3.	<i>Lenses</i>	24	4	3
4.	<i>Tic Tac Toc</i>	958	9	2

5.2 Hasil Uji

Seperti yang telah dibahas pada bab III, terdapat 2 macam pengujian yang dilakukan. Pengujian yang pertama dilakukan untuk mengetahui nilai *purity* pada *New K-Modes* yaitu *K-Modes* yang menggunakan *New Dissimilarity Measure* dan *OLD K-Modes* yaitu *K-Modes* yang Konvensional. Sedangkan pengujian yang kedua dilakukan untuk mengetahui nilai *F-Measure* pada *New K-Modes* yaitu *K-Modes* yang menggunakan *New Dissimilarity Measure* dan *OLD K-Modes* yaitu *K-Modes* yang Konvensional.

5.3 Pengujian untuk mengetahui Nilai *Purity* pada *K-Modes* yang menggunakan *new dissimilarity measure* dan *K-Modes* Konvensional

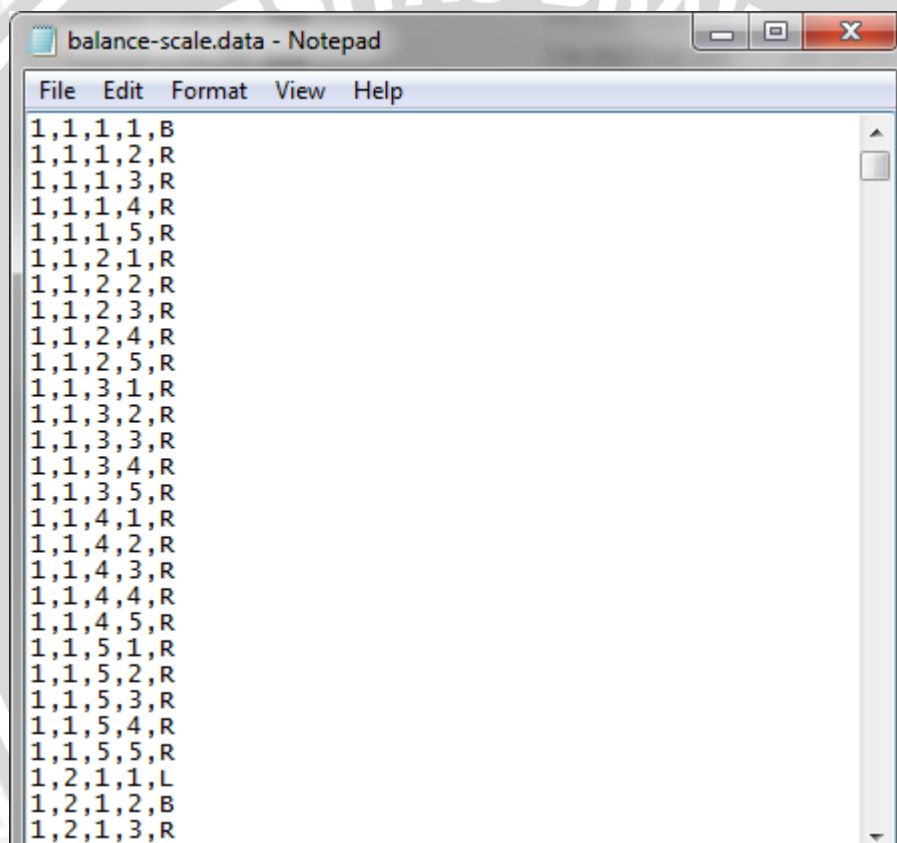
Pengujian yang digunakan untuk mengetahui nilai *purity* pada *K-Modes* menggunakan *New Dissimilarity Measure* dan *K-Modes* Konvensional dilakukan

pada empat *dataset*, yaitu pengujian pertama pada *dataset Balance-Scale*, *Car Evaluation*, *Lenses*, dan *Tic Tac Toe*.

1. *Balance Scale*

Balance Scale memiliki jumlah data sebanyak 625, yang terdiri dari 4 atribut dan *class* didalamnya. Gambar 5.1 adalah contoh dari data yang ada pada *dataset Balance Scale*.

Gambar 5.1 Data *Balance Scale*



Informasi atribut untuk kolom pertama yaitu *Left-Weight* (Berat bagian kiri) yang terdiri dari 5 macam atribut yaitu 1, 2, 3, 4, dan 5. Kolom ke-2 menginformasikan untuk *Left Distance* (Jarak bagian kiri) yang terdiri dari 5 macam atribut yaitu 1, 2, 3, 4, dan 5. Kolom ke-3 menginformasikan *Right-Weight* (Berat bagian kanan) yang terdiri dari 5 macam atribut yaitu 1, 2, 3, 4, dan 5. Kolom ke-4 menginformasikan untuk *Right Distance* (Jarak bagian kanan)

yang terdiri dari 5 macam atribut yaitu 1, 2, 3, 4, dan 5. Sedangkan kolom ke-5 yaitu menginformasikan *class* yang dimiliki oleh data, yang terdiri dari 3 macam yaitu L (*Left*), B (*Balance*), dan R (*Right*).

Seperti yang telah di bahas di bab 1, pengujian yang dilakukan tindak menggunakan data yang memiliki *missing value* didalamnya. Pengujian *clustering* New *K-Modes* pada *dataset Balance Scale* dilakukan pada beberapa macam percobaan. Percobaan pertama menggunakan 50 data dari *dataset Balance Scale* yang ada, ke-2 menggunakan 100 data, ke-3 150 data, ke-4 200 data, ke-5 250 data, ke-6 500 data, dan terakhir menggunakan data keseluruhan yaitu 625 data. Dilakukan 20 pengambilan nilai *purity* pada tiap *dataset*, yang hasilnya berada pada tabel 5.2.

Tabel 5.2 Hasil *Purity* Percobaan New *K-Modes* pada *dataset Balance Scale*

Percobaan Ke-	50 data	100 data	150 data	200 data	250 data	500 data	Full data
1	0.56	0.71	0.72	0.83	0.75	0.68	0.87
2	0.56	0.68	0.85	0.84	0.65	0.77	0.78
3	1	0.68	0.81	0.84	0.57	0.55	0.92
4	1	0.79	0.84	0.83	0.73	0.73	0.88
5	0.56	0.83	0.83	0.77	0.87	0.68	0.85
6	1	0.84	0.60	0.83	0.82	0.69	0.96
7	1	0.66	0.65	0.84	0.62	0.72	0.86
8	1	0.67	0.71	0.79	0.85	0.72	0.81
9	1	0.62	0.85	0.80	0.69	0.69	0.79
10	1	0.76	0.86	0.58	0.88	0.68	0.92
11	1	0.79	0.86	0.86	0.75	0.82	0.89
12	1	0.66	0.86	0.84	0.87	0.67	0.78
13	1	0.68	0.74	0.67	0.84	0.73	0.71
14	1	0.63	0.88	0.83	0.93	0.66	0.85
15	1	0.81	0.85	0.72	0.80	0.80	0.92
16	0.56	0.81	0.63	0.86	0.82	0.67	0.80

17	1	0.59	0.86	0.94	0.76	0.62	0.94
18	1	0.81	0.71	0.83	0.69	0.77	0.77
19	1	0.84	0.87	0.73	0.84	0.62	0.96
20	1	0.66	0.70	0.86	0.83	0.68	0.86
Rata-rata	0.91	0.72	0.78	0.80	0.78	0.70	0.85

Tabel 5.3 berikut menunjukkan nilai *purity* dan *Error* yang dihasilkan pada dataset *Balance Scale* menggunakan *New K-Modes*.

Tabel 5.3 Hasil *Purity* dan *Error* Percobaan *New K-Modes* pada dataset *Balance Scale*

Banyak Data	Nilai <i>Purity</i>	Nilai <i>Error</i>
50 data	0.91	0.09
100 data	0.72	0.28
150 data	0.78	0.22
200 data	0.80	0.20
250 data	0.78	0.22
500 data	0.70	0.30
Full data (625 data)	0.85	0.15
Rata-rata	0.79	0.21

Pengujian *clustering K-Modes* Konvensional pada dataset *Balance Scale* dilakukan pada beberapa macam percobaan. Percobaan pertama menggunakan 50 data dari dataset *Balance Scale* yang ada, ke-2 menggunakan 100 data, ke-3 150 data, ke-4 200 data, ke-5 250 data, ke-6 500 data, dan terakhir menggunakan data keseluruhan yaitu 625 data. Dilakukan 20 pengambilan nilai *purity* pada tiap dataset, yang hasilnya berada pada tabel 5.4.

Tabel 5.4 Hasil *Purity* Percobaan *K-Modes* Konvensional pada *dataset Balance Scale*

Percobaan Ke-	50 data	100 data	150 data	200 data	250 data	500 data	Full data
1	0.72	0.80	0.86	0.66	0.70	0.68	0.53
2	0.68	0.67	0.50	0.66	0.59	0.40	0.58
3	0.58	0.58	0.56	0.61	0.48	0.64	0.69
4	0.68	0.61	0.78	0.58	0.61	0.74	0.62
5	0.62	0.86	0.69	0.74	0.39	0.67	0.51
6	0.68	0.74	0.63	0.58	0.2	0.71	0.48
7	0.68	0.83	0.66	0.63	0.54	0.63	0.66
8	0.80	0.69	0.76	0.69	0.63	0.53	0.66
9	0.74	0.76	0.60	0.83	0.82	0.61	0.65
10	0.58	0.62	0.68	0.65	0.54	0.61	0.70
11	0.62	0.70	0.65	0.81	0.68	0.50	0.54
12	0.56	0.83	0.53	0.83	0.81	0.48	0.57
13	0.54	0.71	0.74	0.79	0.47	0.78	0.55
14	0.46	0.60	0.60	0.60	0.40	0.58	0.79
15	0.50	0.74	0.69	0.76	0.47	0.58	0.64
16	0.62	0.87	0.65	0.69	0.50	0.72	0.59
17	0.80	0.56	0.58	0.62	0.61	0.67	0.60
18	0.72	0.50	0.50	0.54	0.72	0.70	0.48
19	0.64	0.75	0.66	0.56	0.60	0.71	0.78
20	0.68	0.62	0.56	0.56	0.43	0.70	0.55
Rata-rata	0.64	0.70	0.64	0.67	0.56	0.63	0.61

Tabel 5.5 berikut menunjukkan nilai *purity* dan *Error* yang dihasilkan pada *dataset Balance Scale* menggunakan *K-Modes* Konvensional.

Tabel 5.5 Hasil *Purity* dan *Error* Percobaan *K-Modes* Konvensional pada *dataset Balance Scale*

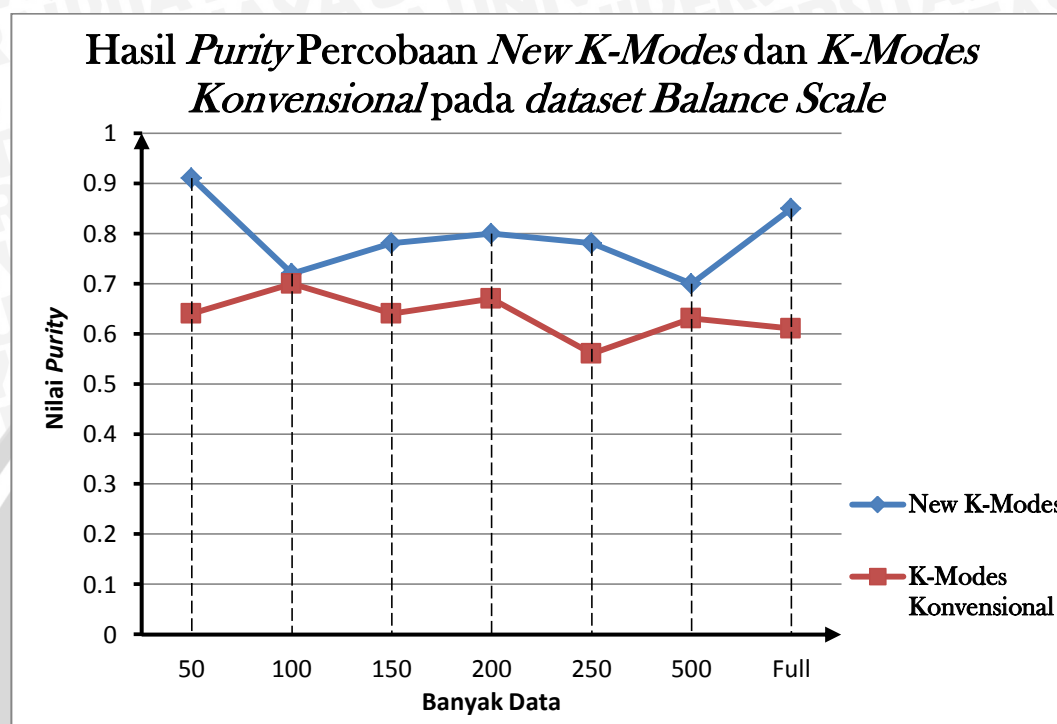
Banyak Data	Nilai <i>Purity</i>	Nilai <i>Error</i>
50 data	0.64	0.36
100 data	0.70	0.30
150 data	0.64	0.36
200 data	0.67	0.33
250 data	0.56	0.44
500 data	0.63	0.37
Full data (625 data)	0.61	0.39
Rata-rata	0.64	0.36

Tabel 5.6 Hasil *Purity* Percobaan *New K-Modes* dan *K-Modes* Konvensional pada *dataset Balance Scale*

Banyak Data	<i>New K-Modes</i>	<i>K-Modes</i> Konvensional
50 data	0.91	0.64
100 data	0.72	0.70
150 data	0.78	0.64
200 data	0.80	0.67
250 data	0.78	0.56
500 data	0.70	0.63
Full data (625 data)	0.85	0.61
Rata-rata	0.79	0.64

Gambar 5.2 merupakan grafik nilai *purity* yang dihasilkan menggunakan *New K-Modes* dan *K-Modes* Konvensional pada *dataset Balance Scale*. Dapat dilihat pada Tabel 5.6, pada percobaan menggunakan 50 data *New K-Modes* didapatkan nilai *purity* 0.91 sedangkan pada *K-Modes* Konvensional 0.64. Pada percobaan menggunakan 100 data, *New K-Modes* didapatkan nilai *purity* 0.72 sedangkan pada *K-Modes* Konvensional 0.70 hingga percobaan menggunakan

keseluruhan data, *New K-Modes* mendapatkan nilai *purity* sebesar 0.85 sedangkan pada *K-Modes Konvensional* mendapatkan 0.61.



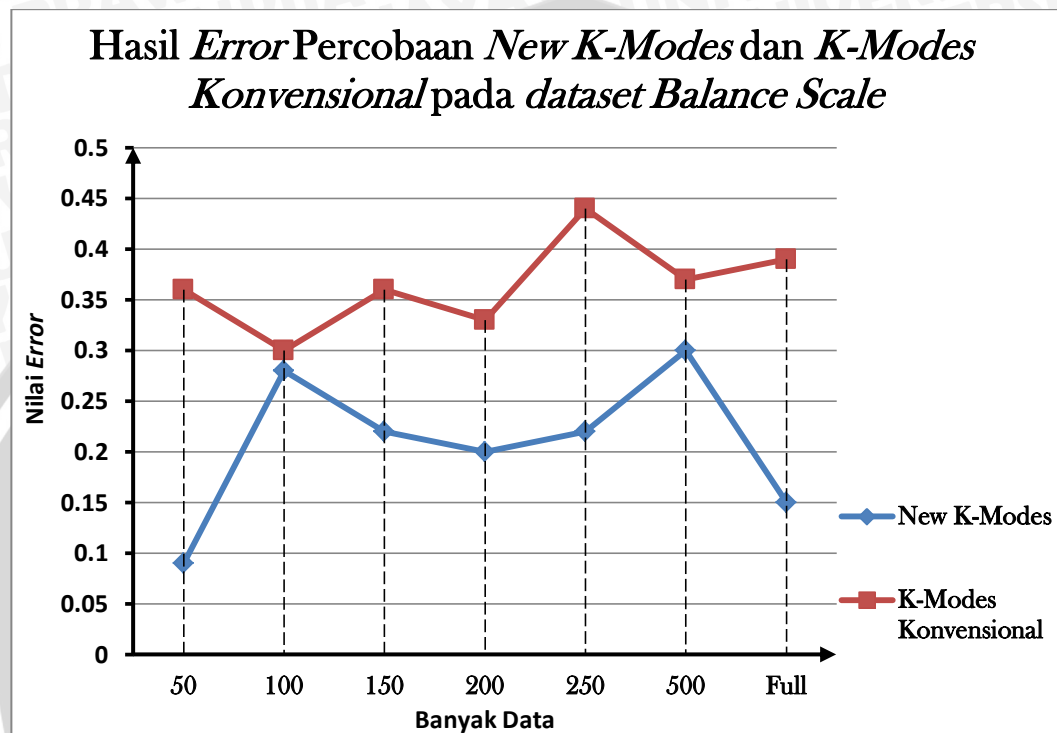
Gambar 5.2 Grafik Hasil *Purity* Percobaan *New K-Modes* dan *K-Modes Konvensional* pada dataset *Balance Scale*

Dilihat dari grafik pada Gambar 5.2 dapat disimpulkan pada percobaan dataset *Balance Scale* menggunakan *New K-Modes* menghasilkan nilai *purity* yang lebih tinggi dibandingkan dengan *K-Modes Konvensional*.

Tabel 5.7 Hasil *Error* Percobaan *New K-Modes* dan *K-Modes Konvensional* pada dataset *Balance Scale*

Banyak Data	<i>New K-Modes</i>	<i>K-Modes Konvensional</i>
50 data	0.09	0.36
100 data	0.28	0.30
150 data	0.22	0.36
200 data	0.20	0.33
250 data	0.22	0.44

500 data	0.30	0.37
Full data (625 data)	0.15	0.39
Rata-rata	0.21	0.36

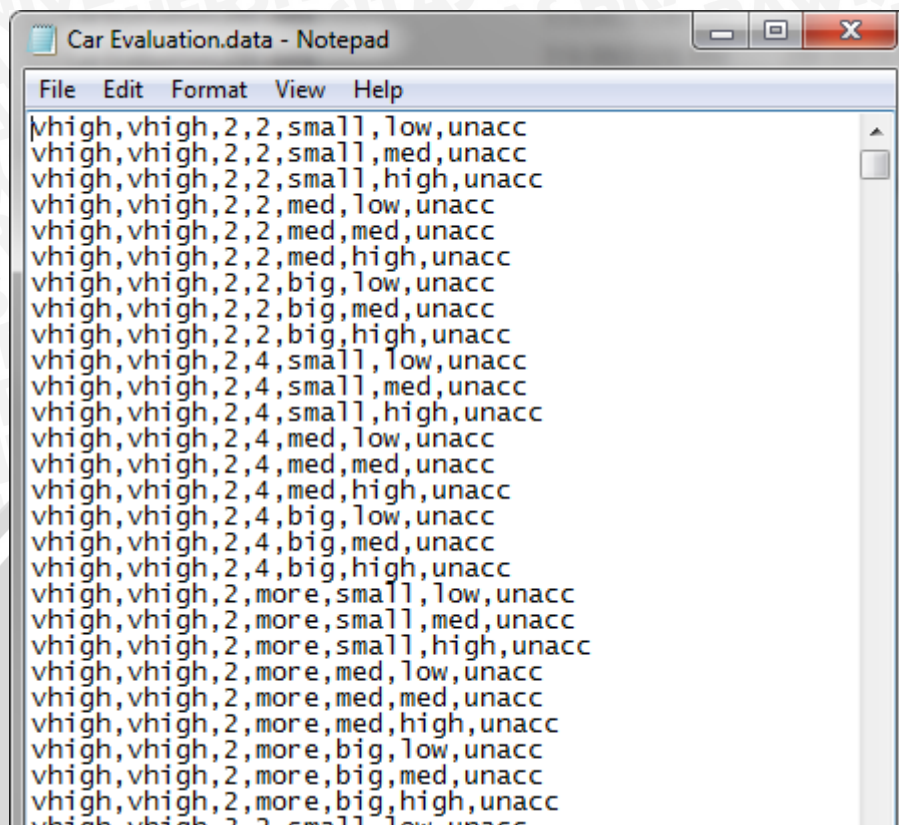


Gambar 5.3 Grafik Hasil *Error* Percobaan *New K-Modes* dan *K-Modes Konvensional* pada dataset *Balance Scale*

Pada percobaan yang dilakukan dataset *Balance Scale* didapatkan nilai *purity* dan *Error* pada Tabel 5.6 dan Tabel 5.7. Dari data yang dihasilkan dapat dilihat pada dataset *Balance sclare* pada $k = 3$, untuk *New K-Modes* memiliki rata-rata nilai *purity* terbaik yaitu 0.79 (79%) dan rata-rata *Error* rendah sebanyak 0.21 (21%). Sedangkan *K-Modes Konvensional* memiliki rata-rata nilai *purity* sebanyak 0.64 (64%) dan rata-rata *Error* sebanyak 0.36 (36%).

2. *Car Evaluation*

Car Evaluation memiliki jumlah data sebanyak 1728, yang terdiri dari 6 attribut dan *class* didalamnya. Berikut Gambar 5.1 adalah contoh dari data yang ada pada dataset *Car Evaluation*.



Gambar 5.4 Data Car Evaluation

Informasi atribut untuk kolom pertama yaitu *buying* (dimana menginformasikan harga beli dari mobil yang bersangkutan) yang terdiri dari 4 macam atribut yaitu *v-high* (sangat mahal), *high* (mahal), *med* (medium), dan *low* (murah). Kolom ke-2 menginformasikan untuk *Maint* (*price of the maintenance*/biaya pemeliharaan) yang terdiri dari 4 macam atribut yaitu *v-high* (sangat mahal), *high* (mahal), *med* (medium), dan *low* (murah). Kolom ke-3 yaitu menginformasikan *doors* (*number of doors*/banyaknya pintu yang ada pada mobil) yang terdiri dari 4 macam atribut yaitu 2, 3, 4, dan 5-*more* (lebih dari 5 pintu). Kolom ke-4 menginformasikan untuk *persons* (kapasitas orang yang dapat di angkut) yang terdiri dari 3 macam atribut yaitu 2, 4, dan *more*.

Kolom ke-5 menginformasikan untuk *Lug_boot* (ukuran dari bagasi) yang terdiri dari 3 macam atribut yaitu *small*, *med*, dan *big*. Kolom ke-6 yaitu menginformasikan *safety* (perkiraan keselamatan mobil) yang terdiri dari 3 macam atribut yaitu *low*, *med*, dan *high*. Sedangkan kolom ke-7 yaitu

menginformasikan *class* yang dimiliki oleh data, yang terdiri dari 4 macam yaitu *unacc*, *acc* (*accept*), *good*, dan *v-good* (*very good*).

Seperti yang telah di bahas di bab 1, bahwa pengujian yang dilakukan tidak menggunakan data yang memiliki *missing value* didalamnya. Pengujian *clustering* New *K-Modes* pada *dataset Car Evaluation* dilakukan pada beberapa macam percobaan. Percobaan pertama menggunakan 50 data dari *dataset Car Evaluation* yang ada, ke-2 menggunakan 100 data, ke-3 150 data, ke-4 200 data, ke-5 250 data, ke-6 500 data, dan terakhir data keseluruhannya yaitu 1728 data. Dilakukan 20 pengambilan nilai *purity* untuk tiap *dataset*, yang hasilnya berada pada tabel 5.8.

Tabel 5.8 Hasil *Purity* Percobaan New *K-Modes* pada *dataset Car Evaluation*

Percobaan Ke-	50 data	100 data	150 data	200 data	250 data	500 data	Full data
1	0.74	0.87	0.55	0.89	0.55	0.57	0.78
2	1	0.88	0.47	0.64	0.75	0.86	0.78
3	1	0.96	0.36	0.88	0.74	0.46	0.78
4	1	0.76	1	0.64	0.60	0.41	0.34
5	1	0.87	0.48	0.89	0.61	0.44	0.45
6	1	0.87	1	0.80	0.62	0.46	0.77
7	1	0.83	0.51	0.64	0.57	0.81	0.87
8	1	0.88	0.63	0.80	0.59	0.59	0.36
9	1	0.67	0.61	0.80	0.63	0.46	0.78
10	0.84	0.87	0.60	0.88	0.67	0.82	0.37
11	0.60	0.83	0.50	0.89	0.56	0.59	0.38
12	0.84	0.78	0.52	0.64	0.78	0.53	0.46
13	0.76	1	0.62	0.89	0.52	0.51	0.37
14	0.74	0.78	0.44	0.80	0.62	0.81	0.42
15	0.76	0.87	0.79	1	0.68	0.44	0.81
16	0.42	0.73	0.50	1	0.65	0.47	0.87
17	1	0.87	0.58	0.88	0.45	0.46	0.36

18	1	0.76	0.57	0.64	0.78	0.95	0.42
19	0.78	0.78	0.50	0.63	0.61	0.43	0.81
20	1	0.67	0.55	1	0.56	0.82	0.40
Rata-rata	0.87	0.82	0.59	0.81	0.63	0.59	0.58

Tabel 5.9 berikut menunjukkan nilai *purity* dan *Error* yang dihasilkan pada dataset *Car Evaluation* menggunakan *New K-Modes*.

Tabel 5.9 Hasil *Purity* dan *Error* Percobaan *New K-Modes* pada dataset *Car Evaluation*

Banyak Data	Nilai <i>Purity</i>	Nilai <i>Error</i>
50 data	0.87	0.13
100 data	0.82	0.18
150 data	0.59	0.41
200 data	0.81	0.19
250 data	0.63	0.37
500 data	0.59	0.41
Full data (1728 data)	0.58	0.42
Rata-rata	0.70	0.30

Pengujian *clustering K-Modes* Konvensional pada dataset *Car Evaluation* dilakukan pada beberapa macam percobaan. Percobaan pertama menggunakan 50 data dari dataset *Car Evaluation* yang ada, ke-2 menggunakan 100 data, ke-3 150 data, ke-4 200 data, ke-5 250 data, ke-6 500 data, dan terakhir menggunakan data keseluruhan yaitu 1728 data. Dilakukan 20 pengambilan nilai *purity* pada tiap dataset, yang hasilnya berada pada tabel 5.10.

Tabel 5.10 Hasil *Purity* Percobaan *K-Modes* Konvensional pada *dataset Car Evaluation*

Percobaan Ke-	50 data	100 data	150 data	200 data	250 data	500 data	Full data
1	0.54	0.49	0.63	0.52	0.65	0.54	0.46
2	0.82	0.49	0.72	0.59	0.52	0.45	0.48
3	0.56	0.56	0.68	0.63	0.57	0.52	0.62
4	0.46	0.58	0.70	0.59	0.49	0.55	0.46
5	0.60	0.66	0.60	0.55	0.44	0.53	0.45
6	0.50	0.55	0.61	0.46	0.5	0.57	0.41
7	0.58	0.62	0.71	0.59	0.43	0.51	0.41
8	0.40	0.66	0.50	0.47	0.47	0.58	0.46
9	0.58	0.63	0.42	0.55	0.53	0.58	0.42
10	0.72	0.52	0.68	0.63	0.48	0.46	0.54
11	0.62	0.46	0.63	0.40	0.51	0.53	0.46
12	0.54	0.60	0.53	0.60	0.53	0.47	0.46
13	0.58	0.55	0.58	0.50	0.61	0.42	0.41
14	0.56	0.49	0.48	0.57	0.45	0.47	0.47
15	0.48	0.64	0.50	0.51	0.46	0.52	0.51
16	0.64	0.59	0.60	0.45	0.55	0.41	0.42
17	0.42	0.59	0.55	0.51	0.45	0.36	0.42
18	0.78	0.75	0.55	0.51	0.45	0.52	0.52
19	0.44	0.42	0.44	0.52	0.67	0.53	0.41
20	0.48	0.75	0.54	0.48	0.52	0.53	0.57
Rata-rata	0.56	0.58	0.58	0.53	0.51	0.50	0.47

Tabel 5.11 berikut menunjukkan nilai *purity* dan *Error* yang dihasilkan pada *dataset Car Evaluation* menggunakan *K-Modes* Konvensional.

Tabel 5.11 Hasil *Purity* dan *Error* Percobaan *K-Modes* Konvensional pada dataset *Car Evaluation*

Banyak Data	Nilai <i>Purity</i>	Nilai <i>Error</i>
50 data	0.56	0.44
100 data	0.58	0.42
150 data	0.58	0.42
200 data	0.53	0.47
250 data	0.51	0.49
500 data	0.50	0.50
Full data (1728 data)	0.47	0.53
Rata-rata	0.53	0.47

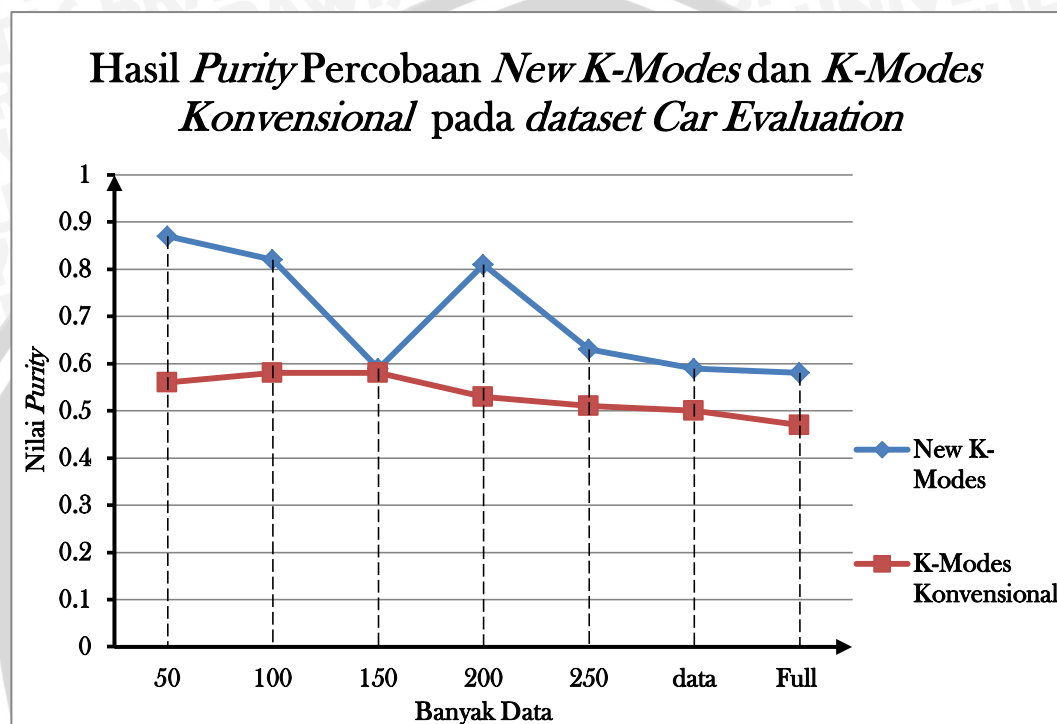
Tabel 5.12 berikut menunjukkan nilai *purity* dan *Error* yang dihasilkan pada dataset *Car Evaluation* menggunakan *New K-Modes* dan *K-Modes* Konvensional.

Tabel 5.12 Hasil *Purity* Percobaan *New K-Modes* dan *K-Modes* Konvensional pada dataset *Car Evaluation*

Banyak Data	<i>New K-Modes</i>	<i>K-Modes</i> Konvensional
50 data	0.87	0.56
100 data	0.82	0.58
150 data	0.59	0.58
200 data	0.81	0.53
250 data	0.63	0.51
500 data	0.59	0.50
Full data (1728 data)	0.58	0.47
Rata-rata	0.70	0.53

Gambar 5.5 merupakan grafik nilai *purity* yang dihasilkan menggunakan *New K-Modes* dan *K-Modes* Konvensional pada dataset *Car Evaluation*. Dapat dilihat pada Tabel 5.12, pada percobaan menggunakan 50 data *New K-Modes* didapatkan nilai *purity* 0.87 sedangkan pada *K-Modes* Konvensional 0.56. Pada

percobaan menggunakan 100 data, *New K-Modes* didapatkan nilai *purity* 0.82 sedangkan pada *K-Modes* Konvensional 0.58 hingga percobaan menggunakan data keseluruhan, *New K-Modes* mendapatkan nilai *purity* 0.58 sedangkan pada *K-Modes* Konvensional mendapatkan 0.47.



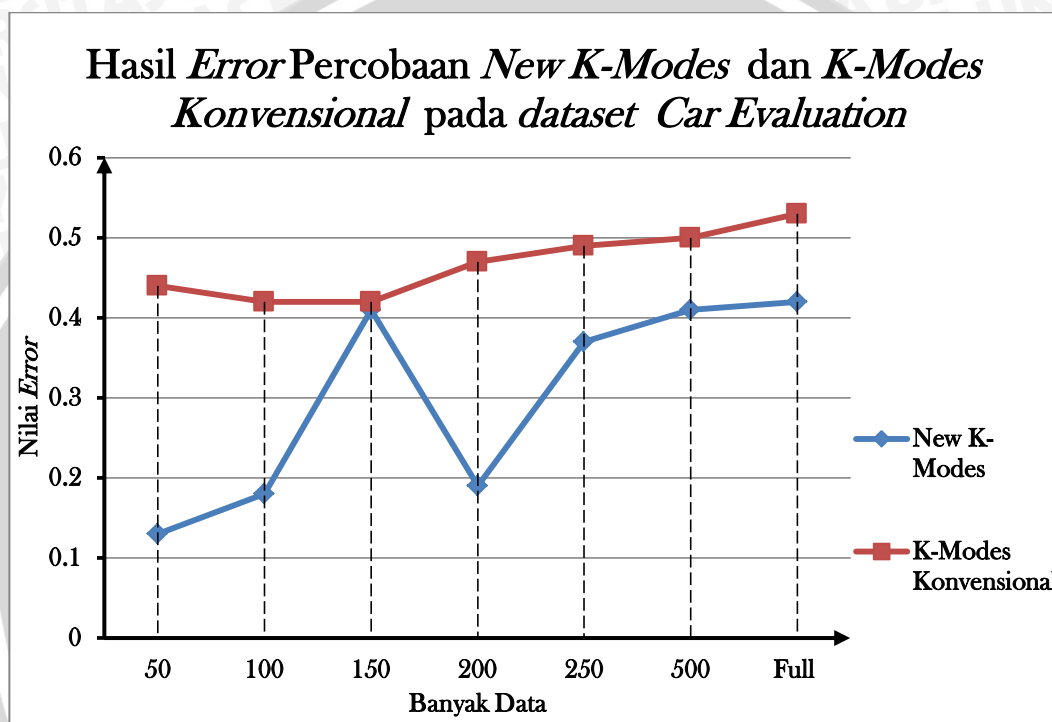
Gambar 5.5 Grafik Hasil *Purity* Percobaan *New K-Modes* dan *K-Modes* Konvensional pada dataset *Car Evaluation*

Dari grafik di Gambar 5.5 dapat disimpulkan pada percobaan dataset *Car Evaluation* menggunakan *New K-Modes* menghasilkan nilai *purity* yang lebih tinggi dibandingkan dengan *K-Modes* Konvensional.

Tabel 5.13 Hasil *Error* Percobaan *New K-Modes* dan *K-Modes* Konvensional pada dataset *Car Evaluation*

Banyak Data	<i>New K-Modes</i>	<i>K-Modes</i> Konvensional
50 data	0.13	0.44
100 data	0.18	0.42
150 data	0.41	0.42

200 data	0.19	0.47
250 data	0.37	0.49
500 data	0.41	0.50
Full data (1728 data)	0.42	0.53
Rata-rata	0.30	0.47



Gambar 5.6 Grafik Hasil *Error* Percobaan *New K-Modes* dan *K-Modes Konvensional* pada dataset *Car Evaluation*

Pada percobaan yang dilakukan dataset *Car Evaluation* didapatkan nilai *purity* dan *Error* pada Tabel 5.12 dan Tabel 5.13. Dari data yang dihasilkan dapat dilihat pada dataset *Car Evaluation* pada $k = 4$, untuk *New K-Modes* memiliki rata-rata nilai *purity* sebanyak 0.70 (70%) dan rata-rata *Error* sebanyak 0.30 (30%). Sedangkan *K-Modes Konvensional* memiliki rata-rata nilai *purity* sebanyak 0.53 (53%) dan rata-rata *Error* sebanyak 0.47 (47%).

3. Lenses

Lenses memiliki jumlah data sebanyak 24, yang terdiri dari 4 atribut dan *class* didalamnya. Berikut Gambar 5.7 adalah contoh dari data yang ada pada *dataset Lenses*.

	File	Edit	Format	View
1	1	1	1	3
1	1	1	2	2
1	1	2	1	3
1	1	2	2	1
1	2	1	1	3
1	2	1	2	2
1	2	2	1	3
1	2	2	2	1
2	1	1	1	3
2	1	1	2	2
2	1	2	1	3
2	1	2	2	1
2	2	1	1	3
2	2	1	2	2
2	2	2	1	3
2	2	2	2	3
3	1	1	1	3
3	1	1	2	3
3	1	2	1	3
3	1	2	2	1
3	2	1	1	3
3	2	1	2	2
3	2	2	1	3
3	2	2	2	3

Gambar 5.7 Data *Lenses*

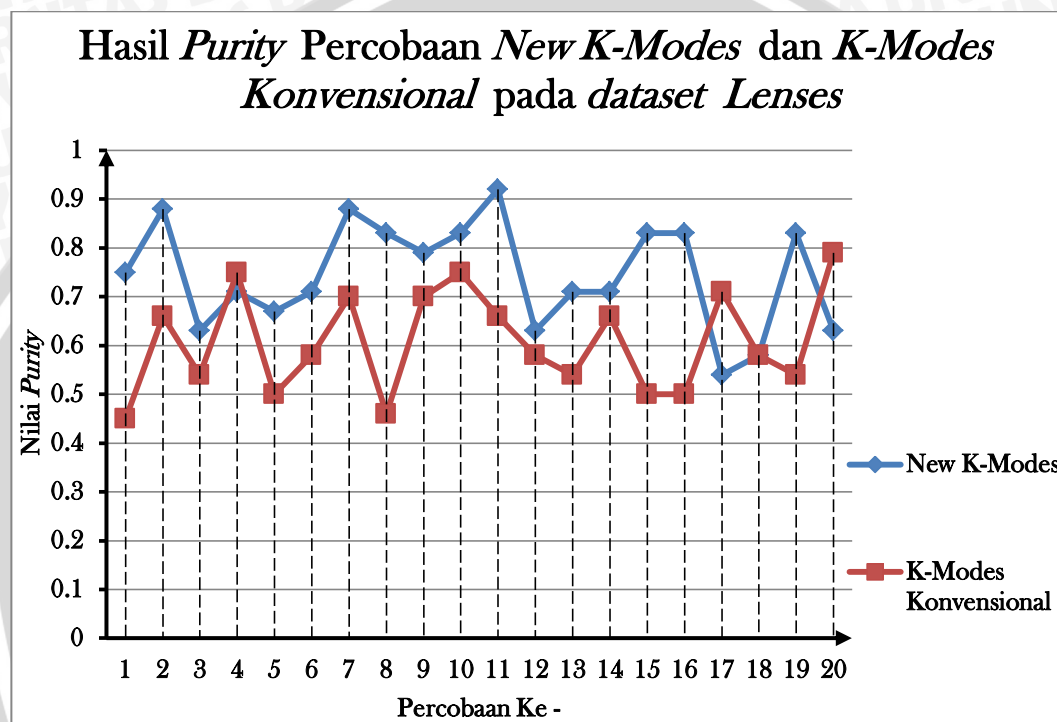
Informasi atribut untuk kolom pertama yaitu umur dari pasien yang terdiri dari 3 macam atribut yaitu *young*, *pre-presbyopic*, dan *presbyopic*. Kolom ke-2 menginformasikan untuk *spectacle prescription* yang terdiri dari 2 macam atribut yaitu *myope*, dan *hypermetrope*. Kolom ke-3 yaitu menginformasikan *astigmatic* yang terdiri dari 2 macam atribut yaitu *no*, dan *yes*. Kolom ke-4 menginformasikan untuk *tear production rate* (produksi air mata) yang terdiri dari 2 macam atribut yaitu *reduced*, dan *normal*. Sedangkan kolom ke-5 yaitu menginformasikan *class* yang dimiliki oleh data, yang terdiri dari 3 macam yaitu *hard contact Lenses*, *soft contact Lenses*, dan *no contact Lenses*.

Seperti yang telah di bahas di bab 1, bahwa pengujian yang dilakukan tidak menggunakan data yang memiliki *missing value* didalamnya. Pengujian *clustering* New *K-Modes* dan *K-Modes* Konvensional pada *dataset Lenses* dilakukan pada data keseluruhannya yaitu 24 data. Dilakukan 20 pengambilan nilai *purity* pada *dataset*, yang hasilnya berada pada tabel 5.14.

Tabel 5.14 Hasil *Purity* Percobaan New *K-Modes* dan *K-Modes* Konvensional pada *dataset Lenses*

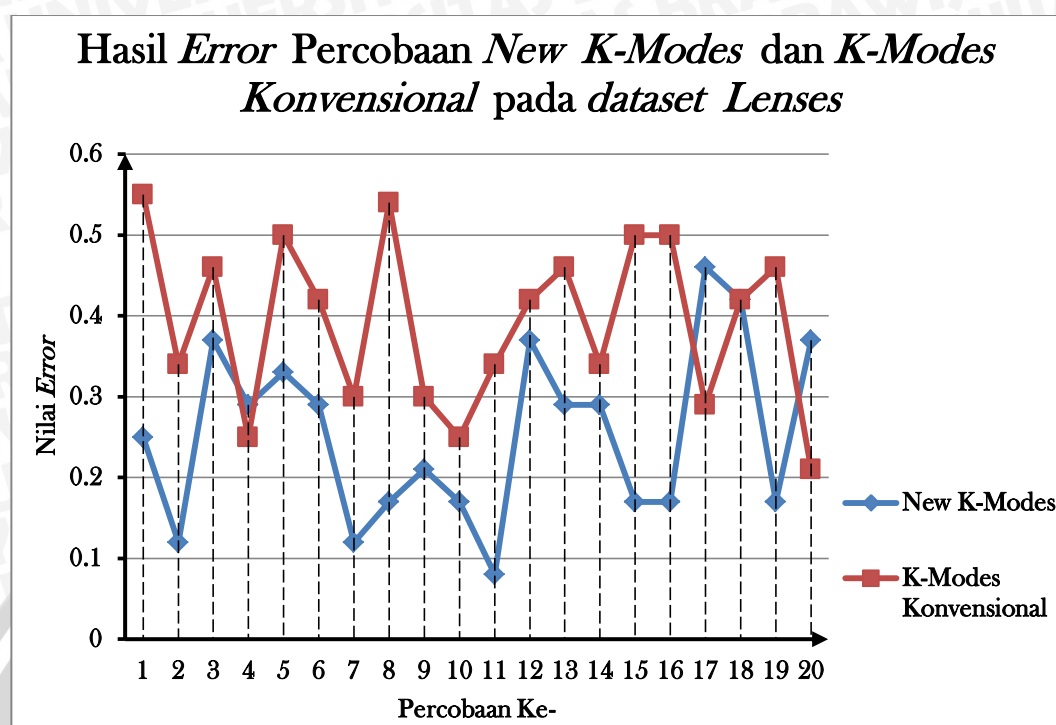
Percobaan Ke -	New <i>K-Modes</i>		<i>K-Modes</i> Konvensional	
	Nilai <i>Purity</i>	Nilai <i>Error</i>	Nilai <i>Purity</i>	Nilai <i>Error</i>
1	0.75	0.25	0.45	0.55
2	0.88	0.12	0.66	0.34
3	0.63	0.37	0.54	0.46
4	0.71	0.29	0.75	0.25
5	0.67	0.33	0.5	0.5
6	0.71	0.29	0.58	0.42
7	0.88	0.12	0.70	0.30
8	0.83	0.17	0.46	0.54
9	0.79	0.21	0.70	0.30
10	0.83	0.17	0.75	0.25
11	0.92	0.08	0.66	0.34
12	0.63	0.37	0.58	0.42
13	0.71	0.29	0.54	0.46
14	0.71	0.29	0.66	0.34
15	0.83	0.17	0.5	0.5
16	0.83	0.17	0.5	0.5
17	0.54	0.46	0.71	0.29
18	0.58	0.42	0.58	0.42
19	0.83	0.17	0.54	0.46
20	0.63	0.37	0.79	0.21
Rata-rata	0.74	0.26	0.61	0.39

Gambar 5.8 merupakan grafik nilai *purity* yang dihasilkan menggunakan *New K-Modes* dan *K-Modes Konvensional* pada *dataset Lenses*. Dapat dilihat pada Tabel 5.6, pada percobaan menggunakan seluruh data *New K-Modes* didapatkan nilai *purity* 0.74 (74%) sedangkan pada *K-Modes Konvensional* mendapatkan 0.61 (61%).



Gambar 5.8 Grafik Hasil *Purity* Percobaan *New K-Modes* dan *K-Modes Konvensional* pada *dataset Lenses*

Dari grafik di Gambar 5.8 dapat disimpulkan pada percobaan *dataset Lenses* menggunakan *New K-Modes* menghasilkan nilai *purity* yang lebih tinggi dibandingkan dengan *K-Modes Konvensional*.

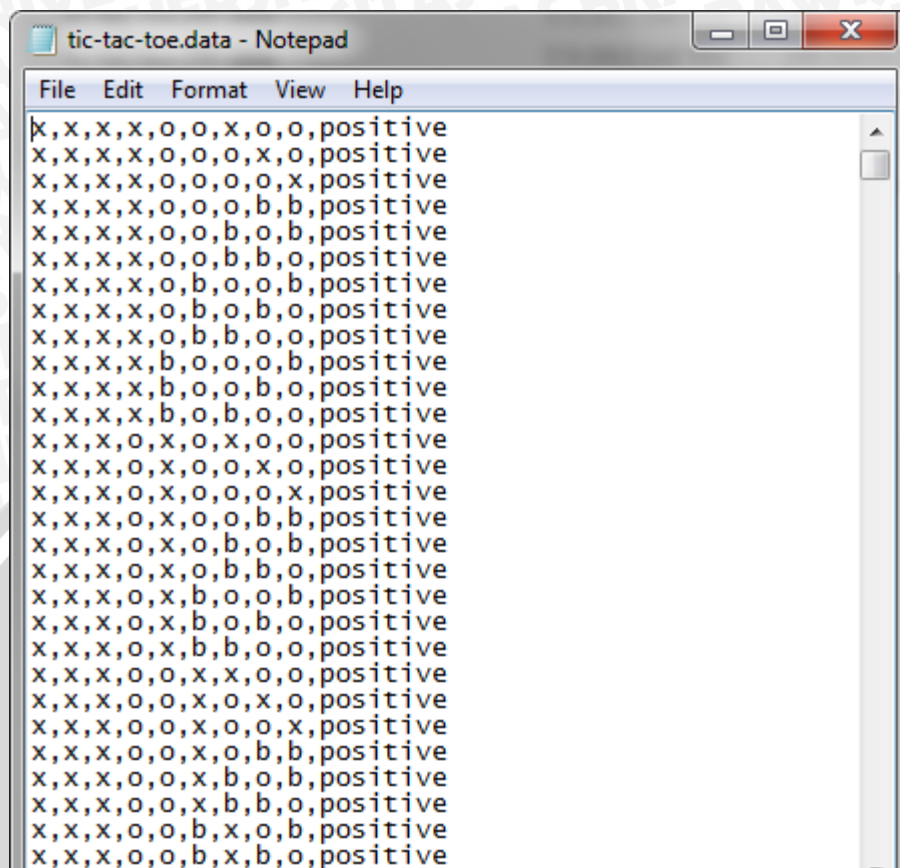


Gambar 5.9 Grafik Hasil *Error* Percobaan *New K-Modes* dan *K-Modes Konvensional* pada dataset *Lenses*

Hasil percobaan yang dilakukan pada dataset *Lenses* didapatkan nilai *purity* dan *Error* pada Tabel 5.14. Dari data yang dihasilkan dapat dilihat pada dataset *Lenses* pada $k = 3$, untuk *New K-Modes* memiliki rata-rata nilai *purity* terbaik sebanyak 0.74 (74%) dan rata-rata *Error* sebanyak 0.26 (26%). Sedangkan *K-Modes Konvensional* memiliki rata-rata nilai *purity* sebanyak 0.61 (61%) dan rata-rata *Error* sebanyak 0.39 (39%).

4. Tic Tac Toc

Tic Tac Toe memiliki jumlah data sebanyak 958, yang terdiri dari 9 atribut dan *class* didalamnya. Gambar 5.10 adalah contoh dari data yang ada pada dataset *Tic Tac Toe*.



Gambar 5.10 Data Tic Tac Toe

Informasi atribut untuk kolom pertama yaitu *top-Left-square* yang terdiri dari 3 macam atribut yaitu x, o dan b. Kolom ke-2 menginformasikan untuk *top-middle-square* yang terdiri dari 3 macam atribut yaitu x, o dan b. Kolom ke-3 yaitu menginformasikan *top-Right-square* yang yang terdiri dari 3 macam atribut yaitu x, o dan b. Kolom ke-4 menginformasikan untuk *middle-Left-square* yang terdiri dari 3 macam atribut yaitu x, o dan b. Kolom ke-5 menginformasikan untuk *middle-middle-square* yang terdiri dari 3 macam atribut yaitu x, o dan b.

Kolom ke-6 yaitu menginformasikan *middle-Right-square* yang yang terdiri dari 3 macam atribut yaitu x, o dan b. Kolom ke-7 menginformasikan untuk *bottom-Left-square* yang terdiri dari 3 macam atribut yaitu x, o dan b. Kolom ke-8 menginformasikan untuk *bottom-middle-square* yang terdiri dari 3 macam atribut yaitu x, o dan b. Kolom ke-9 yaitu menginformasikan *bottom-Right-square* yang yang terdiri dari 3 macam atribut yaitu x, o dan b. Sedangkan kolom ke-10 yaitu

menginformasikan *class* yang dimiliki oleh data, yang terdiri dari 2 macam yaitu *positive*, dan *negative*.

Seperti yang telah di bahas di bab 1, bahwa pengujian yang dilakukan tindak menggunakan data yang memiliki *missing value* didalamnya. Pengujian *clustering* New *K-Modes* pada *dataset Tic Tac Toe* dilakukan pada beberapa macam percobaan. Percobaan pertama menggunakan 50 data dari *dataset Tic Tac Toe* yang ada, ke-2 menggunakan 100 data, ke-3 150 data, ke-4 200 data, ke-5 250 data, ke-6 500 data, dan terakhir menggunakan data keseluruhannya yaitu 958 data. Dilakukan 20 pengambilan nilai *purity* pada tiap *dataset*, yang hasilnya berada pada tabel 5.15.

Tabel 5.15 Hasil *Purity* Percobaan New *K-Modes* pada *dataset Tic Tac Toe*

Percobaan Ke-	50 data	100 data	150 data	200 data	250 data	500 data	Full data
1	0.74	1	1	1	1	1	0.82
2	0.94	0.98	1	1	1	1	0.60
3	1	1	1	1	1	1	0.72
4	1	1	1	1	1	1	0.82
5	0.74	1	1	1	1	1	0.97
6	0.74	1	1	1	0.66	1	0.71
7	0.74	1	1	1	1	1	0.72
8	0.74	1	1	0.60	1	1	0.71
9	0.94	1	1	0.69	1	0.61	0.76
10	0.74	1	1	1	1	0.61	0.71
11	0.74	1	1	1	0.56	1	1
12	0.74	0.85	1	1	0.56	1	0.77
13	1	1	1	1	1	1	1
14	0.74	1	1	1	1	1	0.72
15	1	1	1	1	1	1	1
16	0.94	1	1	1	1	1	0.80
17	1	1	1	0.60	1	1	0.72

Percobaan Ke-	50 data	100 data	150 data	200 data	250 data	500 data	Full data
18	0.74	1	1	0.60	1	1	0.71
19	0.74	1	1	1	1	1	0.72
20	0.94	0.89	1	1	1	1	0.97
Rata-rata	0.84	0.98	1	0.92	0.94	0.96	0.80

Tabel 5.16 berikut menunjukkan nilai *purity* dan *Error* yang dihasilkan pada dataset *Tic Tac Toe* menggunakan *New K-Modes*.

Tabel 5.16 Hasil *Purity* dan *Error* Percobaan *New K-Modes* pada dataset *Tic Tac Toe*

Banyak Data	Nilai <i>Purity</i>	Nilai <i>Error</i>
50 data	0.84	0.16
100 data	0.98	0.02
150 data	1	0
200 data	0.92	0.08
250 data	0.94	0.06
500 data	0.96	0.04
Full data (958 data)	0.80	0.2
Rata-rata	0.92	0.08

Pengujian *clustering K-Modes* Konvensional pada dataset *Tic Tac Toe* dilakukan pada beberapa macam percobaan. Percobaan pertama menggunakan 50 data dari dataset *Tic Tac Toe* yang ada, ke-2 menggunakan 100 data, ke-3 150 data, ke-4 200 data, ke-5 250 data, ke-6 500 data, dan terakhir data keseluruhannya yaitu 625 data. Dilakukan 20 pengambilan nilai *purity* pada tiap dataset, yang hasilnya berada pada tabel 5.17.

Tabel 5.17 Hasil *Purity* Percobaan *K-Modes* Konvensional pada dataset *Tic Tac Toe*

Percobaan Ke-	50 data	100 data	150 data	200 data	250 data	500 data	Full data
1	0.62	0.78	0.78	0.70	0.61	0.62	0.63
2	0.64	0.99	0.77	0.64	0.57	0.56	0.57
3	0.62	0.90	0.88	0.53	0.78	0.72	0.54
4	0.58	0.56	0.91	0.73	0.67	0.67	0.80
5	0.62	0.58	0.52	0.64	0.53	0.71	0.75
6	0.72	1	0.87	0.91	0.65	0.59	0.63
7	0.88	0.96	0.90	0.93	0.55	0.67	0.59
8	0.76	0.95	0.82	0.65	0.86	0.75	0.57
9	0.70	0.63	0.61	0.66	0.65	0.61	0.60
10	0.76	0.96	0.65	0.89	0.57	0.53	0.72
11	0.72	0.85	0.54	0.95	0.55	0.76	0.61
12	0.62	0.69	0.73	0.54	0.79	0.63	0.72
13	0.70	0.54	0.60	0.67	0.59	0.62	0.74
14	0.84	0.98	0.66	0.84	0.93	0.62	0.66
15	0.56	0.62	0.88	0.79	0.93	0.78	0.62
16	0.66	0.95	0.61	0.90	0.66	0.57	0.73
17	0.78	0.66	0.76	0.93	0.75	0.66	0.71
18	0.80	0.56	0.82	0.93	0.92	0.64	0.58
19	0.78	1	0.92	0.70	0.83	0.81	0.74
20	0.64	0.59	0.52	0.80	0.71	0.54	0.59
Rata-rata	0.70	0.79	0.74	0.76	0.70	0.65	0.65

Tabel 5.18 berikut menunjukkan nilai *purity* dan *Error* yang dihasilkan pada dataset *Tic Tac Toe* menggunakan *K-Modes* Konvensional.

Tabel 5.18 Hasil *Purity* dan *Error* Percobaan *K-Modes* Konvensional pada dataset *Tic Tac Toe*

Banyak Data	Nilai <i>Purity</i>	Nilai <i>Error</i>
50 data	0.70	0.30
100 data	0.79	0.21
150 data	0.74	0.26
200 data	0.76	0.24
250 data	0.70	0.30
500 data	0.65	0.35
Full data (958 data)	0.65	0.35
Rata-rata	0.71	0.29

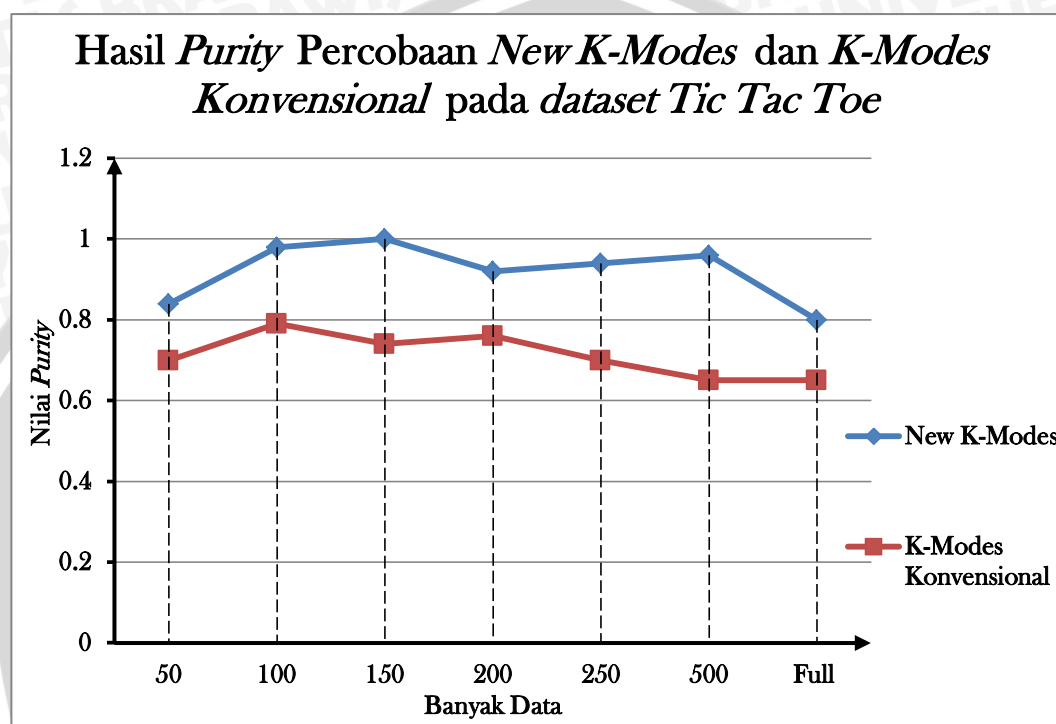
Tabel 5.19 berikut menunjukkan nilai *purity* dan *Error* yang dihasilkan pada dataset *Tic Tac Toe* menggunakan *New K-Modes* dan *K-Modes* Konvensional.

Tabel 5.19 Hasil *Purity* Percobaan *New K-Modes* dan *K-Modes* Konvensional pada dataset *Tic Tac Toe*

Banyak Data	<i>New K-Modes</i>	<i>K-Modes</i> Konvensional
50 data	0.84	0.70
100 data	0.98	0.79
150 data	1	0.74
200 data	0.92	0.76
250 data	0.94	0.70
500 data	0.96	0.65
Full data (958 data)	0.80	0.65
Rata-rata	0.92	0.71

Gambar 5.11 merupakan grafik nilai *purity* yang dihasilkan menggunakan *New K-Modes* dan *K-Modes* Konvensional pada dataset *Tic Tac Toe*. Dapat dilihat pada Tabel 5.15, pada percobaan menggunakan 50 data *New K-Modes* didapatkan nilai *purity* 0.84 sedangkan pada *K-Modes* Konvensional 0.70. Pada

percobaan menggunakan 100 data, *New K-Modes* didapatkan nilai *purity* 0.98 sedangkan pada *K-Modes Konvensional* 0.79 hingga percobaan menggunakan data keseluruhan, *New K-Modes* mendapatkan nilai *purity* 0.80 sedangkan pada *K-Modes Konvensional* mendapatkan 0.65.



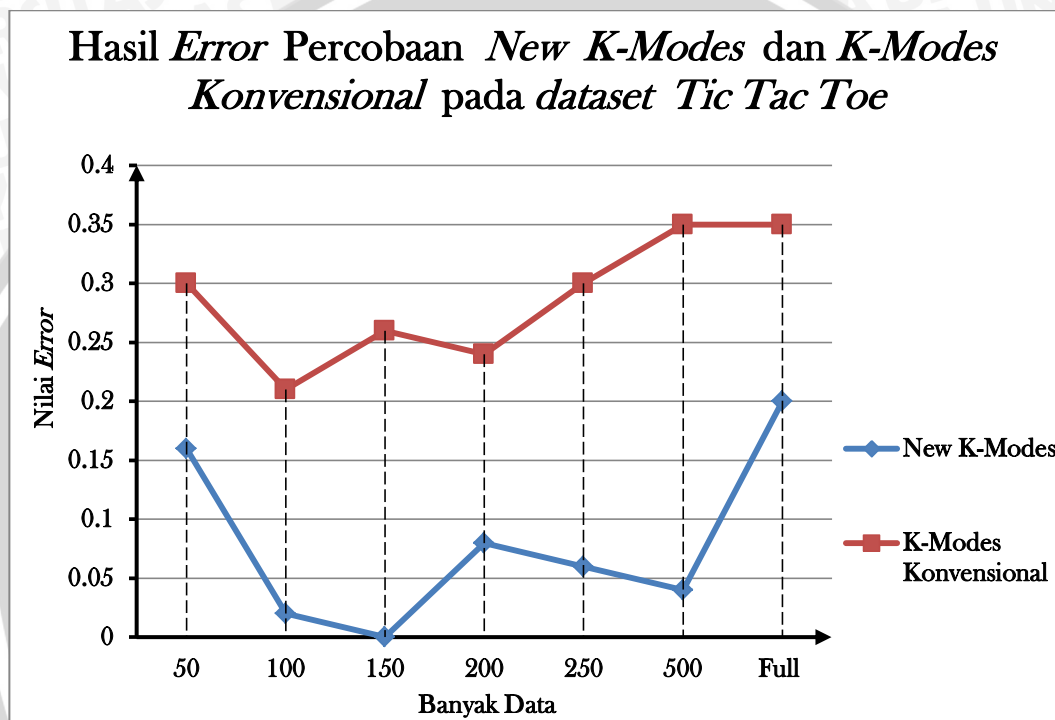
Gambar 5.11 Grafik Hasil *Purity* Percobaan *New K-Modes* dan *K-Modes Konvensional* pada dataset *Tic Tac Toe*

Dari grafik pada Gambar 5.11 dapat disimpulkan pada percobaan dataset *Tic Tac Toe* menggunakan *New K-Modes* menghasilkan nilai *purity* yang lebih tinggi dibandingkan dengan *K-Modes Konvensional*.

Tabel 5.20 Hasil *Error* Percobaan *New K-Modes* dan *K-Modes Konvensional* pada dataset *Tic Tac Toe*

Banyak Data	<i>New K-Modes</i>	<i>K-Modes Konvensional</i>
50 data	0.16	0.30
100 data	0.02	0.21
150 data	0	0.26

200 data	0.08	0.24
250 data	0.06	0.30
500 data	0.04	0.35
Full data (958 data)	0.2	0.35
Rata-rata	0.08	0.29

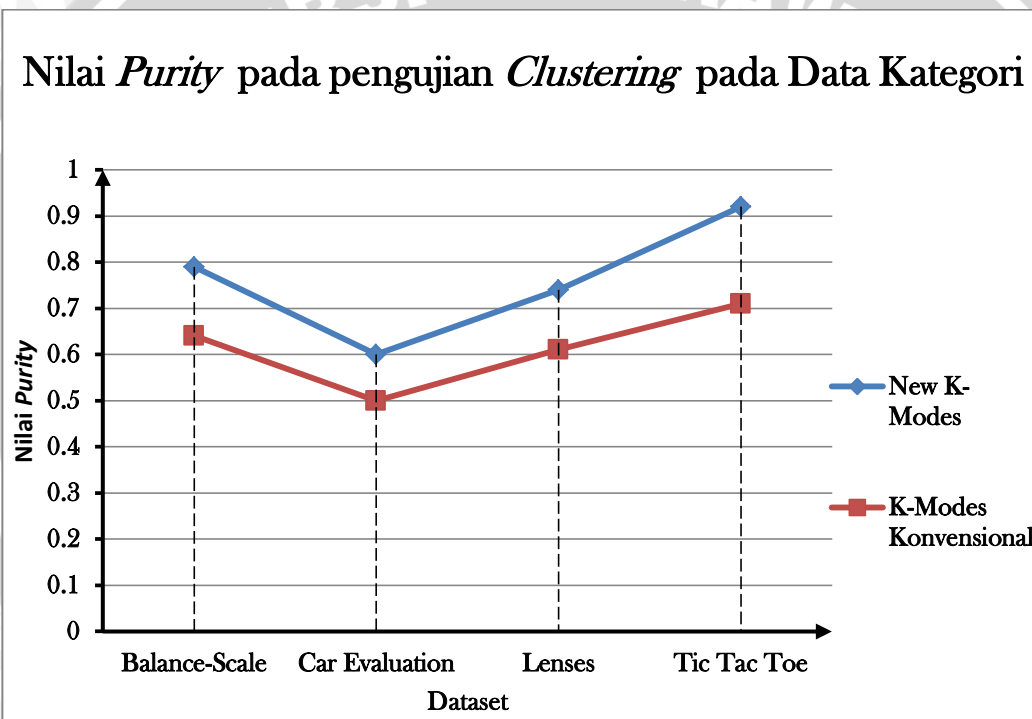


Gambar 5.12 Grafik Hasil *Error* Percobaan *New K-Modes* dan *K-Modes Konvensional* pada dataset *Tic Tac Toe*

Pada percobaan yang dilakukan dataset *Tic Tac Toe* didapatkan nilai *purity* dan *Error* pada Tabel 5.19 dan Tabel 5.20. Dari data yang dihasilkan dapat dilihat pada dataset *Tic Tac Toe* pada $k = 2$, untuk *New K-Modes* memiliki rata-rata nilai *purity* sebanyak 0.92 (92%) dan rata-rata *Error* sebanyak 0.08 (8%). Sedangkan *K-Modes Konvensional* memiliki rata-rata nilai *purity* sebanyak 0.71 (71%) dan rata-rata *Error* sebanyak 0.29 (29%).

Tabel 5.21 Nilai *Purity* pada pengujian *Clustering* pada Data Kategori

<i>Dataset</i>	Nilai <i>Purity</i>	
	<i>New K-Modes</i>	<i>K-Modes Konvensional</i>
<i>Balance-Scale</i>	0.79	0.64
<i>Car Evaluation</i>	0.60	0.50
<i>Lenses</i>	0.74	0.61
<i>Tic Tac Toe</i>	0.92	0.71
Rata-rata	0.76	0.61

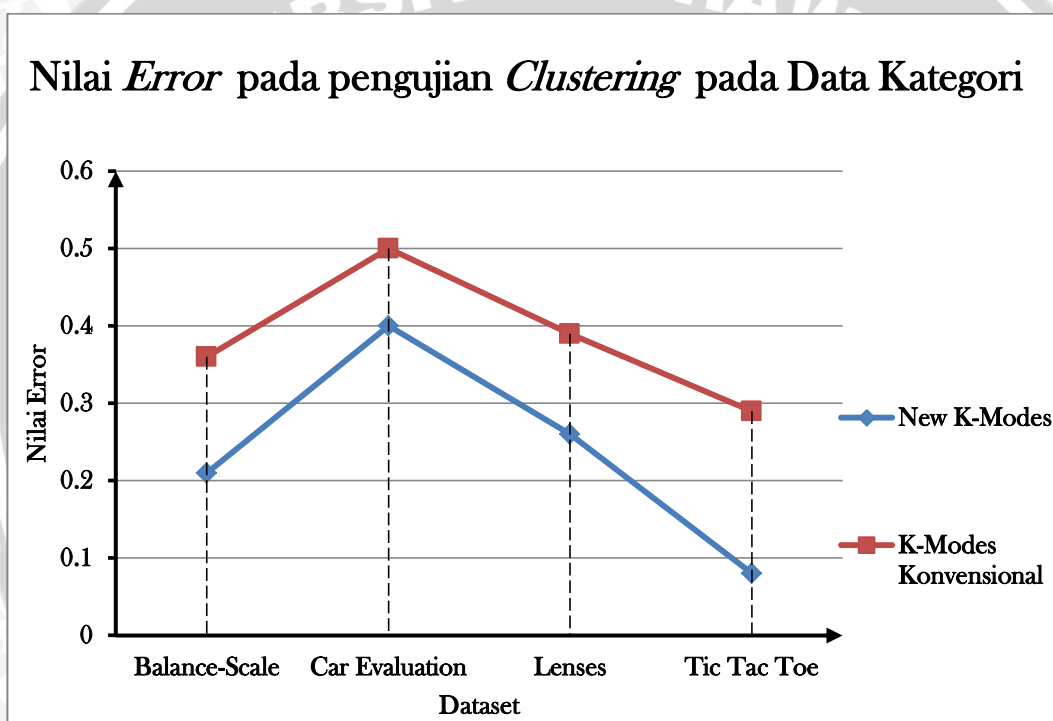


Gambar 5.13 Grafik Nilai *Purity* pada pengujian *Clustering* pada Data Kategori

Gambar 5.13 merupakan grafik hasil nilai *purity* yang dihasilkan pada pengujian *clustering* beberapa data kategori yang menggunakan *New K-Modes* dan *K-Modes Konvensional*. Sedangkan hasil nilai *Error* dari keempat hasil pengujian *dataset* tersebut ditampilkan pada Tabel 5.22.

Tabel 5.22 Error pada pengujian *Clustering* pada Data Kategori

<i>Dataset</i>	Nilai Error	
	<i>New K-Modes</i>	<i>K-Modes Konvensional</i>
<i>Balance-Scale</i>	0.21	0.36
<i>Car Evaluation</i>	0.40	0.50
<i>Lenses</i>	0.26	0.39
<i>Tic Tac Toe</i>	0.08	0.29
Rata-rata	0.21	0.38



Gambar 5.14 Grafik Nilai Error pada pengujian *Clustering* pada Data Kategori

Dari keempat hasil pengujian tersebut didapatkan nilai akurasi atau nilai *Purity K-Modes New Dissimilarity Measure* pada *dataset* tersebut sebesar 0,79; 0,60; 0,74; dan 0,92. Didapatkan nilai akurasi atau nilai *Purity K-Modes Konvensional* pada *dataset* tersebut sebesar 0,64; 0,50; 0,61; dan 0,71. Dimana pada pengujian *purity* menggunakan *New K-Modes* didapatkan rata-rata nilai *purity* sebesar 76% sedangkan rata-rata nilai *purity* pada *K-Modes Konvensional* sebesar 61% yang ditunjukkan pada Tabel 5.21.

5.3.1 Analisis hasil pengujian

Tabel 5.23 Pengujian *Clustering* pada Data Kategori

Dataset	Attribut	Class	Nilai <i>Purity</i>	
			New <i>K-Modes</i>	<i>K-Modes</i> Konvensional
Balance-Scale (625 data)	4	3	0.79	0.64
Car Evaluation (1728 data)	6	4	0.60	0.50
Lenses (24 data)	4	3	0.74	0.61
Tic Tac Toe (958 data)	9	2	0.92	0.71
Rata-rata			0.76	0.61

Pada Tabel 5.23 dapat dilihat bahwa pada pengujian keempat *dataset* yang ada, *dataset* Tic Tac Toe yang mendapatkan nilai *purity* terbaik menggunakan *new K-Modes* yaitu 0.92. Meskipun Tic Tac Toe memiliki data yang banyak (958 data) tetapi *dataset* Tic Tac Toe hanya memiliki 2 *class* saja sehingga membuat peluang nilai *purity* yang dihasilkan semakin kuat.

Dari seluruh hasil percobaan *K-Modes New Dissimilarity Measure* memiliki rata-rata nilai *purity* sebesar 76% sedangkan pada *K-Modes Konvensional* sebesar 61%. Berdasarkan hasil pengujian data kategori menggunakan *K-Modes New Dissimilarity Measure* dan *K-Modes Konvensional* terhadap nilai akurasi atau hasil rating nilai *purity*, sehingga dapat disimpulkan bahwa pengklasteran data kategori menggunakan *New Dissimilarity Measure* memiliki hasil *cluster* yang lebih murni (*pure*) atau memiliki nilai akurasi *cluster* yang lebih baik dan memiliki nilai *Error* yang lebih kecil dibandingkan dengan *clustering* pada *K-Modes Konvensional*.

5.4 Pengujian untuk mengetahui Nilai *F-Measure* pada *K-Modes* yang menggunakan *new dissimilarity measure* dan *K-Modes* Kovenisional

1. Evaluasi *F-Measure* pada dataset *Balance Scale*

Tabel 5.24 Pengujian *Precision* pada Data *Balance Scale* New *K-Modes*

Percobaan Ke-	50 data	100 data	150 data	200 data	250 data	500 data	Full (625 data)
1	0.767	0.532	0.852	0.844	0.711	0.553	0.915
2	0.756	0.891	0.715	0.854	0.875	0.499	0.899
3	0.881	0.533	0.706	0.573	0.898	0.89	0.553
4	0.832	0.8	0.654	0.841	0.859	0.904	0.901
5	0.607	0.576	0.878	0.795	0.745	0.892	0.909
6	0.744	0.895	0.753	0.622	0.828	0.832	0.532
7	0.829	0.807	0.908	0.839	0.877	0.907	0.875
8	0.926	0.759	0.871	0.662	0.574	0.885	0.901
9	0.539	0.763	0.817	0.619	0.847	0.579	0.901
10	1	0.905	0.884	0.857	0.813	0.888	0.589
Rata-rata	0.7881	0.7461	0.8038	0.7506	0.8027	0.7829	0.7975

Tabel 5.25 Pengujian *Recall* pada Data *Balance Scale* New *K-Modes*

Percobaan Ke-	50 data	100 data	150 data	200 data	250 data	500 data	Full (625 data)
1	0.825	0.41	0.87	0.912	0.787	0.43	0.998
2	0.754	0.992	1	0.947	0.98	0.447	0.986
3	1	0.833	0.636	1	0.992	0.969	0.612
4	0.832	0.841	1	0.924	0.933	1	0.977
5	0.559	0.615	0.944	0.842	0.776	0.979	0.999
6	0.756	1	0.712	1	0.961	0.97	0.526
7	0.963	0.791	0.991	0.906	0.984	0.995	0.987
8	1	0.833	0.935	0.679	1	0.996	0.987
9	1	0.737	0.856	0.713	0.894	0.442	0.988

10	1	1	1	0.747	0.878	0.982	0.491
Rata-rata	0.8689	0.8052	0.8944	0.867	0.9185	0.821	0.8551

Tabel 5.26 Pengujian *F-Measure* pada Data *Balance Scale New K-Modes*

Percobaan	50	100	150	200	250	500	Full
Ke-	data	data	data	data	data	data	(625 data)
1	0.753	0.429	0.848	0.867	0.727	0.454	0.954
2	0.754	0.938	0.823	0.897	0.923	0.466	0.94
3	0.934	0.63	0.646	0.723	0.942	0.927	0.533
4	0.83	0.777	0.783	0.871	0.894	0.949	0.937
5	0.57	0.567	0.91	0.792	0.713	0.933	0.951
6	0.749	0.943	0.718	0.766	0.886	0.891	0.516
7	0.882	0.798	0.947	0.856	0.927	0.949	0.926
8	0.959	0.745	0.9	0.668	0.729	0.936	0.942
9	0.698	0.745	0.828	0.586	0.861	0.468	0.942
10	1	0.949	0.938	0.784	0.839	0.931	0.52
Rata-rata	0.8129	0.7521	0.8341	0.781	0.8441	0.7904	0.8161

Pengujian *clustering New K-Modes* pada *dataset Balance Scale* dilakukan pada beberapa macam percobaan. Percobaan pertama menggunakan 50 data dari *dataset Balance Scale* yang ada, ke-2 menggunakan 100 data, ke-3 150 data, ke-4 200 data, ke-5 250 data, ke-6 500 data, dan terakhir data keseluruhannya yaitu 625 data. Dilakukan 10 pengambilan nilai *F-Measure* pada tiap *dataset*, yang hasilnya berada pada tabel 5.26.

Tabel 5.27 Pengujian *F-Measure* pada Data *Balance Scale New K-Modes*

Banyak Data	Precision	Recall	F-Measure
50	0.7881	0.8689	0.8129
100	0.7461	0.8052	0.7521
150	0.8038	0.8944	0.8341
200	0.7506	0.867	0.781

250	0.8027	0.9185	0.8441
500	0.7829	0.821	0.7904
Full	0.7975	0.8551	0.8161
Rata-rata	0.78167	0.86144	0.80438571

Tabel 5.28 Pengujian *Precision* pada Data *Balance Scale K-Modes Konvensional*

Percobaan Ke-	50 data	100 data	150 data	200 data	250 data	500 data	Full (625 data)
1	0.848	0.776	0.773	0.779	0.686	0.784	0.796
2	0.702	0.784	0.749	0.767	0.634	0.662	0.823
3	0.79	0.748	0.66	0.609	0.842	0.605	0.825
4	0.8	0.677	0.833	0.857	0.761	0.517	0.712
5	0.717	0.49	0.687	0.649	0.634	0.772	0.55
6	0.622	0.798	0.752	0.696	0.841	0.751	0.499
7	0.62	0.79	0.87	0.725	0.554	0.597	0.555
8	0.594	0.579	0.838	0.705	0.836	0.706	0.733
9	0.795	0.644	0.813	0.667	0.579	0.81	0.849
10	0.701	0.654	0.778	0.889	0.753	0.547	0.736
Rata-rata	0.7189	0.694	0.7753	0.7343	0.712	0.6751	0.7078

Tabel 5.29 Pengujian *Recall* pada Data *Balance Scale K-Modes Konvensional*

Percobaan Ke-	50 data	100 data	150 data	200 data	250 data	500 data	Full (625 data)
1	0.828	0.769	1	0.853	0.704	0.804	0.914
2	0.639	0.75	0.861	0.85	0.72	0.693	0.849
3	0.866	0.825	0.908	1	0.923	0.588	0.89
4	0.914	0.555	0.87	0.921	0.792	0.472	0.814
5	0.594	0.407	1	1	0.666	0.763	0.649
6	0.622	0.825	0.766	0.663	0.91	0.882	0.534
7	0.719	0.76	0.906	0.935	1	0.697	0.649

8	0.52	0.481	0.868	0.801	0.772	0.861	0.836
9	0.951	0.638	0.847	0.93	1	0.857	0.886
10	0.788	0.63	0.789	0.952	0.728	0.472	0.814
Rata-rata	0.7441	0.664	0.8815	0.8905	0.8215	0.7089	0.7835

Tabel 5.30 Pengujian *F-Measure* pada Data *Balance Scale K-Modes Konvensional*

Percobaan Ke-	50 data	100 data	150 data	200 data	250 data	500 data	Full (625 data)
1	0.829	0.771	0.862	0.814	0.683	0.788	0.841
2	0.659	0.763	0.746	0.785	0.643	0.666	0.835
3	0.817	0.772	0.723	0.754	0.878	0.595	0.849
4	0.833	0.583	0.85	0.887	0.773	0.493	0.719
5	0.625	0.436	0.808	0.782	0.638	0.767	0.544
6	0.598	0.809	0.757	0.672	0.859	0.775	0.488
7	0.638	0.77	0.886	0.792	0.703	0.59	0.523
8	0.541	0.506	0.852	0.736	0.786	0.728	0.752
9	0.852	0.605	0.828	0.776	0.732	0.816	0.858
10	0.73	0.638	0.779	0.919	0.735	0.501	0.751
Rata-rata	0.7122	0.6653	0.8091	0.7917	0.743	0.6719	0.716

Pengujian *clustering K-Modes Konvensional* pada *dataset Balance Scale* dilakukan pada beberapa macam percobaan. Percobaan pertama menggunakan 50 data dari *dataset Balance Scale* yang ada, ke-2 menggunakan 100 data, ke-3 150 data, ke-4 200 data, ke-5 250 data, ke-6 500 data, dan terakhir data keseluruhannya yaitu 625 data. Dilakukan 10 pengambilan nilai *F-Measure* pada tiap *dataset*, yang hasilnya berada pada tabel 5.30.

Tabel 5.31 Pengujian Rata-rata *F-Measure* pada Data *Balance Scale K-Modes Konvensional*

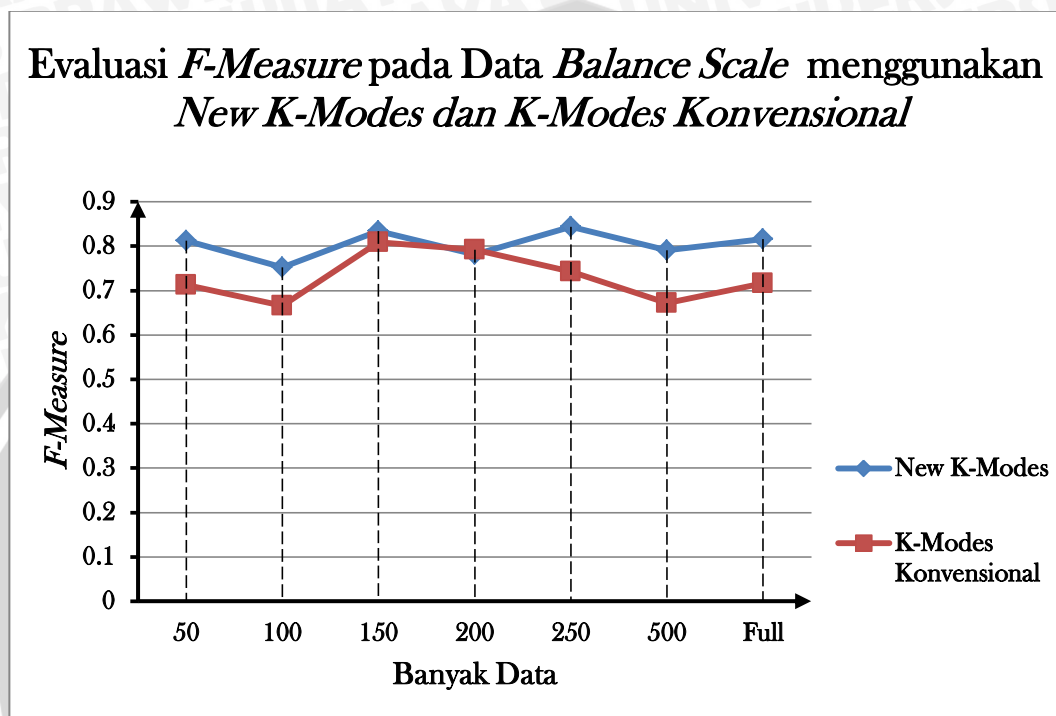
Banyak Data	Precision	Recall	F-Measure
50	0.7189	0.7441	0.7122
100	0.694	0.664	0.6653
150	0.7753	0.8815	0.8091
200	0.7343	0.8905	0.7917
250	0.712	0.8215	0.743
500	0.6751	0.7089	0.6719
Full	0.7078	0.7835	0.716
Rata-rata	0.71677	0.78486	0.72989

Tabel 5.32 Evaluasi *F-Measure* pada Data *Balance Scale* menggunakan *New K-Modes dan K-Modes Konvensional*

Banyak Data	<i>F-Measure</i>	
	<i>New K-Modes</i>	<i>K-Modes Konvensional</i>
50	0.8129	0.7122
100	0.7521	0.6653
150	0.8341	0.8091
200	0.781	0.7917
250	0.8441	0.743
500	0.7904	0.6719
Full	0.8161	0.716
Rata-rata	0.80438571	0.72989

Gambar 5.15 merupakan grafik evaluasi *F-Measure* pada data *balance scale* menggunakan *New K-Modes dan K-Modes Konvensional*. Dapat dilihat pada Tabel 5.32, pada percobaan menggunakan 50 data *New K-Modes* didapatkan nilai *F-Measure* 0.8129 sedangkan pada *K-Modes Konvensional* 0.7122. Pada percobaan menggunakan 100 data, *New K-Modes* didapatkan nilai *F-Measure* 0.7521 sedangkan pada *K-Modes Konvensional* 0.6653 hingga percobaan menggunakan data keseluruhan, *New K-Modes* mendapatkan nilai *F-Measure*

0.8161 sedangkan pada *K-Modes* Konvensional mendapatkan 0.716. Maka di dapatkan rata-rata nilai *F-Measure* *New K-Modes* didapatkan nilai **0.80438571** sedangkan pada *K-Modes* Konvensional **0.72989**.



Gambar 5.15 Grafik Evaluasi *F-Measure* pada Data *Balance Scale* menggunakan *New K-Modes* dan *K-Modes* Konvensional

Dari grafik pada Gambar 5.15 dapat disimpulkan pada percobaan *dataset Balance Scale* menggunakan *New K-Modes* rata-rata menghasilkan nilai *F-Measure* yang lebih tinggi dibandingkan dengan *K-Modes* Konvensional.

2. Evaluasi *F-Measure* pada *dataset Car Evaluation*

Tabel 5.33 Pengujian *Precision* pada Data *Car Evaluation Scale* *New K-Modes*

Percobaan	250	500	750	1000	1300	1500	Full
Ke-	data	data	data	data	data	data	(1728 data)
1	0.688	0.678	0.78	0.757	0.688	0.872	0.796
2	0.641	0.746	0.716	0.743	0.899	0.72	0.831

3	0.714	0.682	0.717	0.729	0.819	0.724	0.72
4	0.631	0.625	0.711	0.713	0.739	0.841	0.834
5	0.702	0.733	0.712	0.761	0.804	0.854	0.689
6	0.762	0.548	0.743	0.705	0.882	0.738	0.824
7	0.632	0.649	0.704	0.711	0.85	0.814	0.71
8	0.699	0.727	0.677	0.685	0.816	0.711	0.839
9	0.696	0.759	0.719	0.613	0.88	0.873	0.805
10	0.715	0.567	0.706	0.803	0.856	0.834	0.815
Rata-rata	0.688	0.6714	0.7185	0.722	0.8233	0.7981	0.7863

Tabel 5.34 Pengujian *Recall* pada Data *Car Evaluation Scale New K-Modes*

Percobaan Ke-	250 data	500 data	750 data	1000 data	1300 data	1500 data	Full (1728 data)
1	1	0.826	0.828	0.84	0.666	0.82	0.874
2	0.711	0.878	1	0.76	0.925	1	0.891
3	0.783	0.862	1	0.789	0.814	1	1
4	0.79	0.674	1	0.833	1	0.88	0.895
5	0.889	0.917	1	0.788	0.833	0.77	1
6	0.884	0.583	1	0.699	0.855	1	0.882
7	0.489	0.944	1	0.894	0.825	0.773	1
8	0.787	0.976	1	0.809	0.868	1	0.904
9	0.755	0.799	1	1	0.886	0.89	0.874
10	0.791	0.581	1	0.898	0.862	0.861	0.861
Rata-rata	0.7879	0.804	0.9828	0.831	0.8534	0.8994	0.9181

Tabel 5.35 Pengujian *F-Measure* pada Data *Car Evaluation Scale New K-Modes*

Percobaan Ke-	250 data	500 data	750 data	1000 data	1300 data	1500 data	Full (1728 data)
1	0.81	0.743	0.791	0.787	0.666	0.84	0.825
2	0.643	0.785	0.832	0.742	0.91	0.831	0.852

3	0.733	0.747	0.835	0.754	0.805	0.834	0.83
4	0.691	0.642	0.831	0.762	0.846	0.845	0.86
5	0.776	0.801	0.831	0.765	0.809	0.804	0.813
6	0.811	0.555	0.851	0.696	0.867	0.842	0.838
7	0.525	0.765	0.825	0.765	0.837	0.792	0.826
8	0.739	0.808	0.806	0.734	0.819	0.831	0.865
9	0.681	0.76	0.834	0.755	0.882	0.874	0.827
10	0.743	0.57	0.827	0.847	0.842	0.825	0.817
Rata-rata	0.7152	0.7176	0.8263	0.7607	0.8283	0.8318	0.8353

Pengujian *clustering New K-Modes* pada *dataset Car Evaluation* dilakukan pada beberapa macam percobaan. Percobaan pertama menggunakan 250 data dari *dataset Car Evaluation* yang ada, ke-2 menggunakan 500 data, ke-3 750 data, ke-4 1000 data, ke-5 1300 data, ke-6 1500 data, dan terakhir data keseluruhannya yaitu 1728 data. Dilakukan 10 pengambilan nilai *F-Measure* pada tiap *dataset*, yang hasilnya berada pada tabel 5.35.

Tabel 5.36 Pengujian Rata-rata *F-Measure* pada Data *Car Evaluation New K-Modes*

Banyak Data	Precision	Recall	F-Measure
250	0.688	0.7879	0.7152
500	0.6714	0.804	0.7176
750	0.7185	0.9828	0.8263
1000	0.722	0.831	0.7607
1300	0.8233	0.8534	0.8283
1500	0.7981	0.8994	0.8318
Full	0.7863	0.9181	0.8353
Rata-rata	0.74394	0.86809	0.78789

Tabel 5.37 Pengujian *Precision* pada Data *Car Evaluation K-Modes Konvensional*

Percobaan Ke-	250 data	500 data	750 data	1000 data	1300 data	1500 data	Full (1728 data)
1	0.708	0.638	0.744	0.649	0.803	0.857	0.747
2	0.727	0.639	0.754	0.755	0.815	0.823	0.768
3	0.655	0.646	0.774	0.715	0.724	0.847	0.684
4	0.745	0.686	0.786	0.739	0.748	0.723	0.817
5	0.716	0.711	0.715	0.67	0.781	0.822	0.693
6	0.623	0.664	0.756	0.695	0.781	0.779	0.678
7	0.711	0.72	0.803	0.591	0.706	0.686	0.746
8	0.684	0.671	0.722	0.732	0.755	0.817	0.774
9	0.68	0.736	0.71	0.603	0.839	0.827	0.612
10	0.618	0.688	0.853	0.681	0.796	0.801	0.739
Rata-rata	0.6867	0.6799	0.7617	0.683	0.7748	0.7982	0.7258

Tabel 5.38 Pengujian *Recall* pada Data *Car Evaluation K-Modes Konvensional*

Percobaan Ke-	250 data	500 data	750 data	1000 data	1300 data	1500 data	Full (1728 data)
1	0.796	0.763	0.794	0.59	0.687	0.779	0.748
2	1	0.694	0.809	0.852	0.639	0.811	0.783
3	0.651	0.768	0.822	0.741	0.666	0.832	1
4	0.829	0.706	0.837	0.878	1	1	0.768
5	0.752	0.799	1	0.639	0.718	0.8	1
6	0.646	0.886	0.762	0.644	1	0.77	1
7	0.736	0.785	0.795	1	1	0.55	0.662
8	0.893	0.827	1	0.799	0.795	0.778	0.699
9	1	0.802	1	0.598	0.762	0.788	0.664
10	1	0.712	0.845	0.645	0.754	0.737	1
Rata-rata	0.8303	0.7742	0.8664	0.7386	0.8021	0.7845	0.8324

Tabel 5.39 Pengujian *F-Measure* pada Data *Car Evaluation K-Modes Konvensional*

Percobaan Ke-	250 data	500 data	750 data	1000 data	1300 data	1500 data	Full (1728 data)
1	0.742	0.688	0.747	0.612	0.735	0.812	0.746
2	0.822	0.659	0.741	0.786	0.691	0.809	0.769
3	0.647	0.699	0.785	0.727	0.677	0.833	0.81
4	0.78	0.692	0.796	0.781	0.853	0.837	0.787
5	0.721	0.745	0.834	0.622	0.745	0.804	0.799
6	0.616	0.741	0.759	0.661	0.87	0.757	0.794
7	0.691	0.735	0.799	0.72	0.825	0.604	0.695
8	0.762	0.736	0.838	0.759	0.73	0.793	0.729
9	0.794	0.755	0.826	0.595	0.795	0.802	0.612
10	0.758	0.633	0.849	0.654	0.769	0.763	0.844
Rata-rata	0.7333	0.7083	0.7974	0.6917	0.769	0.7814	0.7585

Pengujian *clustering K-Modes Konvensional* pada *dataset Car Evaluation* dilakukan pada beberapa macam percobaan. Percobaan pertama menggunakan 250 data dari *dataset Car Evaluation* yang ada, ke-2 menggunakan 500 data, ke-3 750 data, ke-4 1000 data, ke-5 1300 data, ke-6 1500 data, dan terakhir data keseluruhannya yaitu 1728 data. Dilakukan 10 pengambilan nilai *F-Measure* pada tiap *dataset*, yang hasilnya berada pada tabel 5.39.

Tabel 5.40 Pengujian Rata-rata *F-Measure* pada Data *Car Evaluation K-Modes Konvensional*

Banyak Data	Precision	Recall	F-Measure
250	0.6867	0.8303	0.7333
500	0.6799	0.7742	0.7083
750	0.7617	0.8664	0.7974
1000	0.683	0.7386	0.6917
1300	0.7748	0.8021	0.769
1500	0.7982	0.7845	0.7814

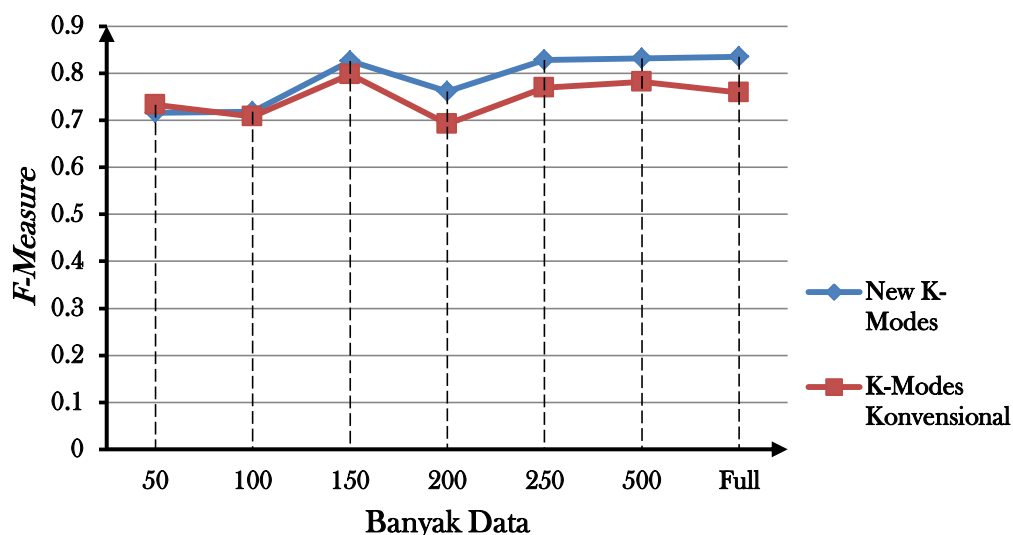
Full	0.7258	0.8324	0.7585
Rata-rata	0.73001	0.80407	0.74851429

Tabel 5.41 Evaluasi *F-Measure* pada Data *Car Evaluation* menggunakan *New K-Modes* dan *K-Modes Konvensional*

Banyak Data	<i>F-Measure</i>	
	<i>New K-Modes</i>	<i>K-Modes Konvensional</i>
250	0.7152	0.7333
500	0.7176	0.7083
750	0.8263	0.7974
1000	0.7607	0.6917
1300	0.8283	0.769
1500	0.8318	0.7814
Full	0.8353	0.7585
Rata-rata	0.78789	0.74851429

Gambar 5.16 merupakan grafik evaluasi *F-Measure* pada data *car evaluation* menggunakan *New K-Modes* dan *K-Modes Konvensional*. Dapat dilihat pada Tabel 5.41, pada percobaan menggunakan 250 data *New K-Modes* didapatkan nilai *F-Measure* 0.7152 sedangkan pada *K-Modes Konvensional* 0.7333. Pada percobaan menggunakan 500 data, *New K-Modes* didapatkan nilai *F-Measure* 0.7176 sedangkan pada *K-Modes Konvensional* 0.7083 hingga percobaan menggunakan data keseluruhan, *New K-Modes* mendapatkan nilai *F-Measure* 0.8353 sedangkan pada *K-Modes Konvensional* mendapatkan 0.7585. Maka di dapatkan rata-rata nilai *F-Measure* *New K-Modes* didapatkan nilai **0.78789** sedangkan pada *K-Modes Konvensional* **0.74851**.

Evaluasi *F-Measure* pada Data *Car Evaluation* menggunakan *New K-Modes* dan *K-Modes Konvensional*



Gambar 5.16 Grafik Evaluasi *F-Measure* pada Data *Car Evaluation* menggunakan *New K-Modes* dan *K-Modes Konvensional*

Dari grafik pada Gambar 5.16 dapat disimpulkan pada percobaan *dataset Car Evaluation* menggunakan *New K-Modes* rata-rata menghasilkan nilai *F-Measure* yang lebih tinggi dibandingkan dengan *K-Modes Konvensional*.

3. Evaluasi *F-Measure* pada *dataset Lenses*

Tabel 5.42 Pengujian *F-Measure* pada Data *Lenses New K-Modes*

Percobaan Ke -	Precision	Recall	F-Measure
1	0.722	1	0.822
2	0.667	1	0.778
3	1	1	1
4	0.833	1	0.889
5	0.833	1	0.889
6	0.833	1	0.889
7	0.722	1	0.822
8	0.833	1	0.889
9	0.667	1	0.778

10	0.889	0.833	0.822
Rata-rata	0.7999	0.9833	0.8578

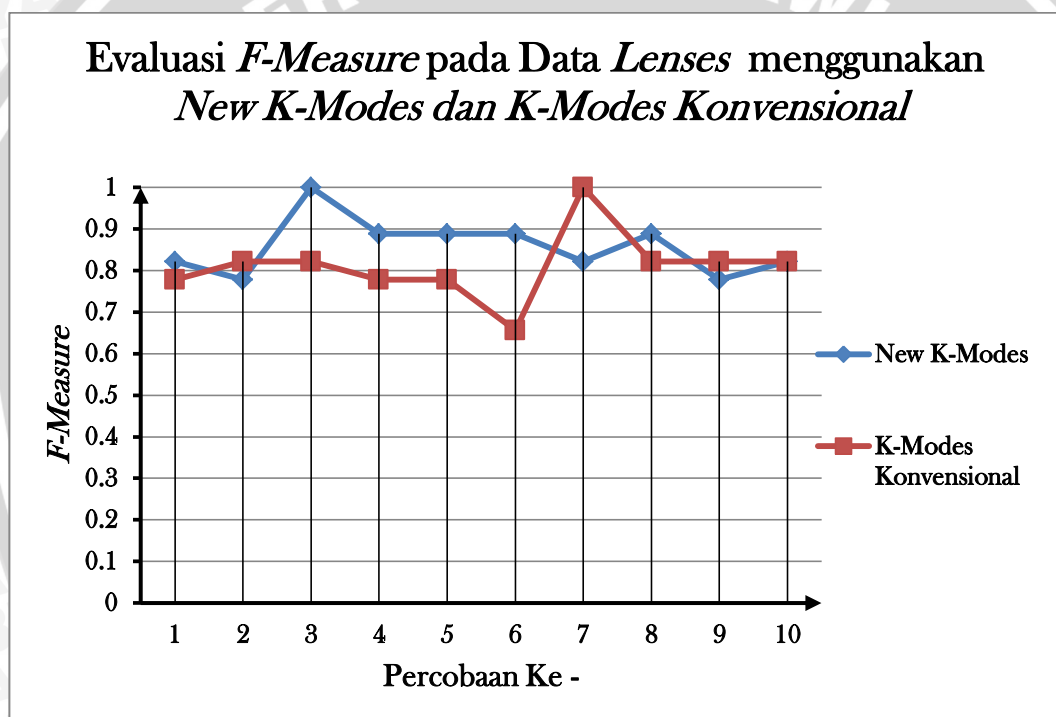
Tabel 5.43 Pengujian *F-Measure* pada Data *Lenses K-Modes Konvensional*

Percobaan Ke -	Precision	Recall	F-Measure
1	0.667	1	0.778
2	0.722	1	0.822
3	0.889	0.833	0.822
4	0.667	1	0.778
5	0.667	1	0.778
6	0.722	0.667	0.656
7	1	1	1
8	0.722	1	0.822
9	0.889	0.833	0.822
10	0.722	1	0.822
Rata-rata	0.7667	0.9333	0.81

Tabel 5.44 Evaluasi *F-Measure* pada Data *Lenses* menggunakan *New K-Modes* dan *K-Modes Konvensional*

Percobaan Ke -	<i>F-Measure</i>	
	<i>New K-Modes</i>	<i>K-Modes Konvensional</i>
1	0.822	0.778
2	0.778	0.822
3	1	0.822
4	0.889	0.778
5	0.889	0.778
6	0.889	0.656
7	0.822	1
8	0.889	0.822
9	0.778	0.822
10	0.822	0.822
Rata-rata	0.8578	0.81

Gambar 5.17 merupakan grafik evaluasi *F-Measure* pada data *lenses* menggunakan *New K-Modes* dan *K-Modes Konvensional*. Dapat dilihat pada Tabel 5.44, pada percobaan pertama *New K-Modes* didapatkan nilai *F-Measure* 0.822 sedangkan pada *K-Modes Konvensional* 0.778. Pada percobaan kedua, *New K-Modes* didapatkan nilai *F-Measure* 0.778 sedangkan pada *K-Modes Konvensional* 0.822 hingga percobaan ke - 10, *New K-Modes* mendapatkan nilai *F-Measure* 0.822 sedangkan pada *K-Modes Konvensional* mendapatkan 0.822. Maka di dapatkan rata-rata nilai *F-Measure* *New K-Modes* sebesar **0.8578**, sedangkan pada *K-Modes Konvensional* **0.81**.



Gambar 5.17 Evaluasi *F-Measure* pada Data *Lenses* menggunakan *New K-Modes* dan *K-Modes Konvensional*

Dari grafik pada Gambar 5.17 dapat disimpulkan pada percobaan *dataset Lenses* menggunakan *New K-Modes* rata-rata menghasilkan nilai *F-Measure* yang lebih tinggi dibandingkan dengan *K-Modes Konvensional*.

4. Evaluasi *F-Measure* pada dataset *Tic Tac Toe*

Tabel 5.45 Pengujian *Precision* pada Data *Tic Tac Toe New K-Modes*

Percobaan Ke-	50 data	100 data	150 data	200 data	250 data	500 data	Full (958 data)
1	0.748	0.981	0.52	0.579	0.669	0.509	0.746
2	0.619	0.582	0.55	0.947	0.541	0.852	0.659
3	0.946	0.946	1	0.751	0.816	0.582	0.64
4	0.767	0.99	0.921	0.97	0.967	0.582	0.83
5	0.857	0.99	0.927	0.824	0.807	0.528	0.659
6	0.821	0.98	0.954	0.931	0.908	0.523	0.654
7	0.917	1	0.693	0.981	0.548	0.565	0.668
8	0.838	0.601	0.611	0.995	0.57	0.862	0.787
9	0.931	1	1	0.862	0.934	0.839	0.66
10	0.704	0.911	0.955	0.54	0.57	0.51	0.753
Rata-rata	0.8148	0.8981	0.8131	0.838	0.733	0.6352	0.7056

Tabel 5.46 Pengujian *Recall* pada Data *Tic Tac Toe New K-Modes*

Percobaan Ke-	50 data	100 data	150 data	200 data	250 data	500 data	Full (958 data)
1	0.7	0.98	0.52	0.566	0.668	0.508	0.758
2	0.6	0.58	0.547	0.946	0.541	0.84	0.647
3	0.94	0.94	1	0.746	0.816	0.582	1
4	0.76	0.99	0.907	0.97	0.964	0.582	0.752
5	0.8	0.99	0.927	0.817	0.804	0.528	1
6	0.72	0.98	0.953	0.93	0.908	0.522	1
7	0.9	1	0.693	0.98	0.548	0.56	0.678
8	0.76	0.6	0.6	0.995	0.569	0.856	0.803
9	0.92	1	1	0.848	0.932	0.77	1
10	0.64	0.9	0.953	0.54	0.569	0.51	0.7
Rata-rata	0.774	0.896	0.81	0.8338	0.7319	0.6258	0.8338

Tabel 5.47 Pengujian *F-Measure* pada Data *Tic Tac Toe New K-Modes*

Percobaan Ke-	50 data	100 data	150 data	200 data	250 data	500 data	Full (958 data)
1	0.685	0.98	0.52	0.55	0.668	0.5	0.751
2	0.583	0.577	0.54	0.945	0.538	0.839	0.651
3	0.94	0.94	1	0.744	0.816	0.581	0.778
4	0.758	0.99	0.906	0.97	0.964	0.582	0.77
5	0.792	0.99	0.927	0.814	0.804	0.526	0.794
6	0.696	0.98	0.953	0.93	0.908	0.518	0.791
7	0.899	1	0.693	0.98	0.548	0.551	0.67
8	0.745	0.599	0.589	0.995	0.566	0.855	0.793
9	0.919	1	1	0.848	0.932	0.758	0.795
10	0.609	0.899	0.953	0.54	0.566	0.51	0.712
Rata-rata	0.7626	0.8955	0.8081	0.8316	0.731	0.622	0.7505

Tabel 5.48 Pengujian Rata-rata *F-Measure* pada Data *Tic Tac Toe New K-Modes*

Banyak Data	Precision	Recall	F-Measure
50	0.8148	0.774	0.7626
100	0.8981	0.896	0.8955
150	0.8131	0.81	0.8081
200	0.838	0.8338	0.8316
250	0.733	0.7319	0.731
500	0.6352	0.6258	0.622
Full	0.7056	0.8338	0.7505
Rata-rata	0.776	0.786	0.771

Tabel 5.49 Pengujian *Precision* pada Data Tic Tac Toe K-Modes Konvensional

Percobaan Ke-	50 data	100 data	150 data	200 data	250 data	500 data	Full (958 data)
1	0.6	0.625	0.569	0.513	0.526	0.551	0.655
2	0.64	0.595	0.914	0.616	0.894	0.875	0.654
3	0.868	0.651	0.908	0.86	0.949	0.52	0.643
4	0.603	0.963	0.861	0.897	0.992	0.562	0.743
5	0.691	0.574	0.523	0.861	0.976	0.589	0.654
6	0.79	0.901	0.507	0.55	0.778	0.676	0.655
7	0.879	0.63	0.84	0.522	0.516	0.722	0.742
8	0.521	0.565	0.514	0.79	0.808	0.598	0.653
9	0.564	0.541	0.567	0.871	0.555	0.725	0.686
10	0.808	0.92	0.54	0.945	0.917	0.566	0.653
Rata-rata	0.6964	0.6965	0.6743	0.7425	0.7911	0.6384	0.6738

Tabel 5.50 Pengujian *Recall* pada Data Tic Tac Toe K-Modes Konvensional

Percobaan Ke-	50 data	100 data	150 data	200 data	250 data	500 data	Full (958 data)
1	0.6	0.62	0.567	1	0.525	0.55	1
2	0.64	0.58	0.907	0.608	0.892	0.874	1
3	0.82	0.65	0.907	0.86	0.948	0.52	1
4	0.6	0.96	0.847	0.878	0.992	0.562	0.768
5	0.68	0.57	0.52	0.852	0.976	0.582	1
6	0.76	0.9	0.507	0.549	0.776	0.676	1
7	0.84	0.63	0.84	0.522	0.516	0.708	0.754
8	0.52	0.56	0.513	0.784	0.804	0.584	1
9	0.56	0.54	0.567	0.87	0.551	0.712	0.705
10	0.8	0.92	0.54	0.939	0.916	0.566	0.652
Rata-rata	0.682	0.693	0.6715	0.7862	0.7896	0.6334	0.8879

Tabel 5.51 Pengujian *F-Measure* pada Data *Tic Tac Toe K-Modes Konvensional*

Percobaan Ke-	50 data	100 data	150 data	200 data	250 data	500 data	Full (958 data)
1	0.6	0.616	0.562	0.678	0.519	0.548	0.791
2	0.64	0.562	0.906	0.599	0.892	0.874	0.79
3	0.814	0.65	0.907	0.86	0.948	0.52	0.781
4	0.597	0.96	0.845	0.878	0.992	0.562	0.732
5	0.675	0.565	0.505	0.849	0.976	0.573	0.787
6	0.754	0.9	0.501	0.547	0.775	0.676	0.791
7	0.836	0.63	0.84	0.518	0.515	0.703	0.696
8	0.513	0.551	0.511	0.783	0.803	0.569	0.781
9	0.554	0.537	0.566	0.87	0.544	0.708	0.679
10	0.799	0.92	0.539	0.94	0.916	0.566	0.653
Rata-rata	0.6782	0.6891	0.6682	0.7522	0.788	0.6299	0.7481

Tabel 5.52 Pengujian Rata-rata *F-Measure* pada Data *Tic Tac Toe K-Modes Konvensional*

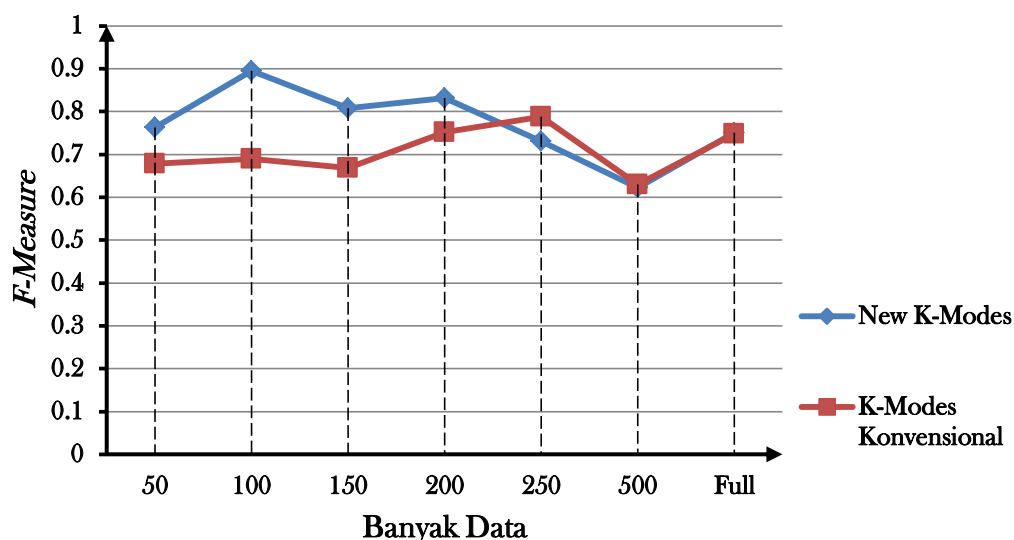
Banyak Data	Precision	Recall	F-Measure
50	0.6964	0.682	0.6782
100	0.6965	0.693	0.6891
150	0.6743	0.6715	0.6682
200	0.7425	0.7862	0.7522
250	0.7911	0.7896	0.788
500	0.6384	0.6334	0.6299
Full	0.6738	0.8879	0.7481
Rata-rata	0.701	0.734	0.707

Tabel 5.53 Evaluasi *F-Measure* pada Data *Tic Tac Toe* menggunakan *New K-Modes* dan *K-Modes Konvensional*

Banyak Data	<i>F-Measure</i>	
	<i>New K-Modes</i>	<i>K-Modes Konvensional</i>
50	0.7626	0.6782
100	0.8955	0.6891
150	0.8081	0.6682
200	0.8316	0.7522
250	0.731	0.788
500	0.622	0.6299
Full	0.7505	0.7481
Rata-rata	0.771	0.707

Gambar 5.18 merupakan grafik evaluasi *F-Measure* pada data *Tic Tac Toe* menggunakan *New K-Modes* dan *K-Modes Konvensional*. Dapat dilihat pada Tabel 5.53, pada percobaan menggunakan 50 data *New K-Modes* didapatkan nilai *F-Measure* 0.7626 sedangkan pada *K-Modes Konvensional* 0.6782. Pada percobaan menggunakan 100 data, *New K-Modes* didapatkan nilai *F-Measure* 0.8955 sedangkan pada *K-Modes Konvensional* 0.6891 hingga percobaan menggunakan data keseluruhan, *New K-Modes* mendapatkan nilai *F-Measure* 0.7505 sedangkan pada *K-Modes Konvensional* mendapatkan 0.7481. Maka di dapatkan rata-rata nilai *F-Measure New K-Modes* sebesar **0.771**, sedangkan pada *K-Modes Konvensional* **0.707**.

Evaluasi *F-Measure* pada Data *Tic Tac Toe* menggunakan *New K-Modes* dan *K-Modes Konvensional*

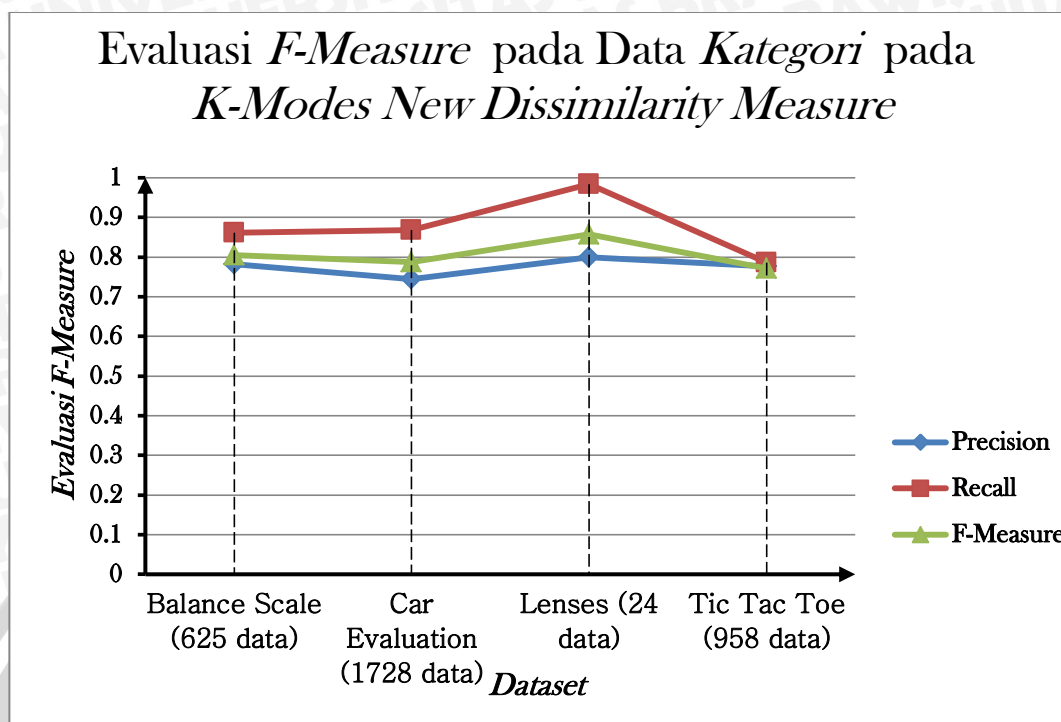


Gambar 5.18 Evaluasi *F-Measure* pada Data *Tic Tac Toe* menggunakan *New K-Modes* dan *K-Modes Konvensional*

Dari grafik pada Gambar 5.18 dapat disimpulkan pada percobaan *dataset Tic Tac Toe* menggunakan *New K-Modes* rata-rata menghasilkan nilai *F-Measure* yang lebih tinggi dibandingkan dengan *K-Modes Konvensional*.

Tabel 5.54 Pengujian Rata-rata *F-Measure* pada Data Kategori pada *K-Modes New Dissimilarity Measure*

<i>Dataset</i>	<i>Class</i>	<i>Attribut</i>	<i>Precision</i>	<i>Recall</i>	<i>F-Measure</i>
<i>Balance Scale (625 data)</i>	3	4	0.78167	0.86144	0.80439
<i>Car Evaluation (1728 data)</i>	4	6	0.74394	0.86809	0.78789
<i>Lenses (24 data)</i>	3	4	0.7999	0.9833	0.8578
<i>Tic Tac Toe (958 data)</i>	2	9	0.77683	0.78647	0.77161
Rata-rata			0.77559	0.87483	0.80542



Gambar 5.19 Evaluasi *F-Measure* pada Data *Kategori* pada *K-Modes New Dissimilarity*

Analisa hasil pengklasteran metode *K-Modes*

Seperti yang terdapat pada Tabel 5.24 hingga Tabel 5.54, maka hasil pengklasteran *K-Modes New Dissimilarity Measure* dapat dianalisa sebagai dalam beberapa aspek, antara lain : analisa pola nilai *precision*, *recall*, dan *F-Measure*, analisa banyak kelas dalam dataset terhadap nilai *precision*, *recall*, dan *F-Measure*.

a. Analisa pola nilai *precision*, *recall*, dan *f-measure*

Pada Tabel 5.24 hingga Tabel 5.54, untuk *dataset* yang sama, berapapun tingkat *accuracy*, dan k = jumlah *class*, nilai *precision*, *recall*, dan *F-Measure* berbeda-beda. Hal ini disebabkan pada k = *class*, nilai *recall* selalu berubah-ubah yang disebabkan adanya nilai *false negative* yang ada pada cluster, sedangkan nilai *true positive* yang tidak lain adalah frekuensi *class* terbanyak dalam *dataset* diikuti oleh nilai *precision*. Dan hasil pengklasteran yang

dilakukan secara *random* dalam menentukan pusat *cluster* membuat nilai *F-Measure* yang dihasilkan tidak konsisten.

b. Analisa banyak kelas dalam dataset terhadap nilai *precision*, *recall*, dan *f-measure*

Pola nilai yang terdapat pada Tabel 5.24 hingga 5.54 berbeda-beda untuk tiap dataset. Pada dataset *balance scale* dengan class sebanyak tiga, kisaran *f-measure* antara 0.75 - 0.84. Untuk dataset *car evaluation* dengan class sebanyak empat, kisaran *f-measure* antara 0,715 - 0.835. Sedangkan pada dataset *lenses* dengan tiga class, kisaran *f-measure* antara 0.778 - 1. Sedangkan pada dataset *tic tac toe* dengan dua class, kisaran *f-measure* antara 0.622 - 0.89. Dari data pada Tabel 5.54 dapat diambil kesimpulan bahwa banyaknya class pada data asli mempengaruhi performa pengklasteran, makin sedikit class dan semakin kecil jumlah data, maka semakin baik performa pengklasteran. Hal ini disebabkan oleh perhitungan *true positive* yang tinggi, sebaliknya *false positive* dan *false negative* akan bernilai rendah untuk tiap cluster. Rata-rata *precision*, *recall*, dan *F-Measure* tiap dataset ditunjukkan Tabel 5.54.

Maka dapat disimpulkan bahwa hasil akurasi penempatan cluster pada dataset data kategori dengan menggunakan *K-Modes New Dissimilarity Measure* memiliki rata-rata nilai *F-Measure* terbaik pada dataset *lenses* sebesar 0.8578, sedangkan rata-rata *F-Measure* dari keseluruhan dataset adalah **0.8054**.

BAB VI

KESIMPULAN DAN SARAN

6.1 Kesimpulan

Dari hasil uji dan analisis yang telah dilakukan, dapat diambil kesimpulan yaitu :

1. Sistem *clustering K-Modes* menggunakan *New Dissimilarity Measure* menggunakan jarak ketidaksamaan dengan mengubah persamaan yang ada di *K-Modes Konvensional* dimana $x = y$ tidak lagi bernilai 0 tetapi $1 - w_{ij}$, dimana merupakan perkalian dari perbandingan atribut yang ada di *cluster* dengan perbandingan nilai atribut di *dataset*.
2. Hasil akurasi penempatan *cluster* pada *dataset* data kategori dengan menggunakan *K-Modes New Dissimilarity Measure* memiliki rata-rata nilai *purity* yang terbaik yaitu sebesar **0.76**, sedangkan rata-rata *purity* pada *K-Modes Konvensional* sebesar **0.61**. Dapat disimpulkan bahwa *K-Modes* menggunakan *New Dissimilarity Measure* memiliki hasil evaluasi yang lebih baik dibandingkan menggunakan *K-Modes Konvensional*. Dengan kata lain semakin besar *purity* yang dihasilkan oleh *K-Modes New Dissimilarity Measure*, membuat *cluster* yang dihasilkan lebih murni (*pure*), karena semua atribut dengan *class* yang sama berada pada *cluster* yang sama.
3. Hasil akurasi penempatan *cluster* pada *dataset* data kategori dengan menggunakan *K-Modes New Dissimilarity Measure* memiliki rata-rata nilai *F-Measure* sebesar , sedangkan rata-rata *F-Measure* adalah **0.8054**.

6.2 Saran

Beberapa saran untuk pengembangan lebih lanjut yang dapat diberikan oleh penulis adalah :

1. Pada umumnya data yang ada banyak yang memiliki missing value (nilai atribut tidak ada). Maka selanjutnya dapat menggunakan *dataset* yang memiliki atribut *missing value*.

2. Penentuan titik pusat awal yang dilakukan secara random, menyebabkan masih ada kesalahan pada penempatan *cluster*. Maka dari itu pada proses *clustering* untuk data kategori selanjutnya bisa dengan menggunakan algoritma *clustering* yang lain sehingga bisa mendapatkan nilai *purity* yang lebih baik lagi.

