

## BAB V PENGUJIAN DAN ANALISIS

Pada bab ini terdapat penjabaran pengujian seperti yang telah dipaparkan pada bab III, yaitu meliputi pengujian nilai  $k$ , pengujian keterlibatan jumlah data, pengujian keterlibatan fitur, serta pengujian jarak. Pada setiap pengujian akan dilakukan sebanyak dua kali dengan jalur SBNMPTN/SNMPTN dan SPMK.

### 5.1 Skenario Pengujian

Terdapat empat skenario dalam pengujian ini, untuk lebih jelasnya akan dijabarkan pada Tabel 5.1 berikut.

Tabel 5.1 Skenario Pengujian

Skenario	Pengujian			
	$k$	Jumlah Data	Fitur	Jarak
1	√			
2		√		
3			√	
4				√

Pada masing-masing skenario terdapat pengujian yang berbeda-beda. Skenario 1 merupakan pengujian akurasi yang dipengaruhi oleh nilai  $k$ , jadi  $k$  berjalan hingga didapat akurasi tertinggi dengan jumlah data latih dan data uji yang konstan. Skenario 2 merupakan pengujian akurasi yang dipengaruhi oleh jumlah data latih dan data uji, jadi komposisi perbandingan jumlah data latih dan data uji berjalan hingga didapat akurasi terbesar dengan nilai  $k$  yang konstan (didapat dari skenario 1).

Skenario 3 merupakan pengujian akurasi yang dipengaruhi oleh keterlibatan fitur saat proses perhitungan. Nilai  $k$  dan jumlah data yang terlibat saat dilakukan skenario 3 adalah konstan, didapat dari hasil perhitungan pada skenario sebelumnya.

Skenario 4 merupakan pengujian akurasi yang dipengaruhi oleh jenis metode perhitungan jarak. Jika saat skenario 1 hingga skenario 3 berjalan dengan metode jarak Euclidean, maka saat skenario 4 dilakukan dengan metode

perhitungan jarak Manhattan. Nilai  $k$ , perbandingan jumlah data serta fitur yang terlibat mengikuti skenario 3.

### 1. Pengujian Skenario 1

Pengujian  $k$  dilakukan dengan cara merubah nilai, mulai  $k=2$  hingga  $k=n$  sampai didapat akurasi maksimum. Pengujian  $k$  dilakukan sebanyak dua kali, yaitu terhadap jalur masuk SBNMPTN/SNMPTN dan SPMK. Perbandingan data latih dan data ujinya sebanyak 80% : 20% dari 1.004 dan 256 data. Hasil pengujian  $k$  ditampilkan dalam Tabel 5.2 berikut :

Tabel 5.2 Hasil Akurasi Skenario 1 Jalur SBNMPTN/SNMPTN

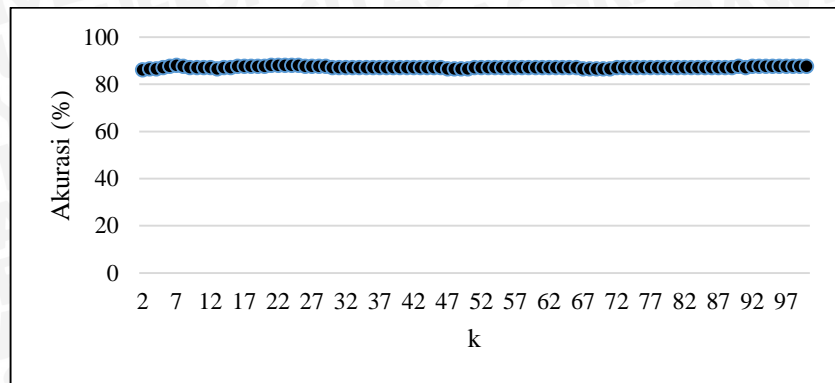
<b>k</b>	<b>Data Latih (%)</b>	<b>Data Uji (%)</b>	<b>Akurasi (%)</b>	<b>k</b>	<b>Data Latih (%)</b>	<b>Data Uji (%)</b>	<b>Akurasi (%)</b>
2	80	20	86.07	53	80	20	87.06
3	80	20	86.57	54	80	20	87.06
4	80	20	86.57	55	80	20	87.06
5	80	20	87.06	56	80	20	87.06
6	80	20	87.56	57	80	20	87.06
7	80	20	88.06	58	80	20	87.06
8	80	20	87.56	59	80	20	87.06
9	80	20	87.06	60	80	20	87.06
10	80	20	87.06	61	80	20	87.06
11	80	20	87.06	62	80	20	87.06
12	80	20	87.06	63	80	20	87.06
13	80	20	86.57	64	80	20	87.06
14	80	20	87.06	65	80	20	87.06
15	80	20	87.06	66	80	20	87.06
16	80	20	87.56	67	80	20	86.57
17	80	20	87.56	68	80	20	86.57
18	80	20	87.56	69	80	20	86.57
19	80	20	87.56	70	80	20	86.57
20	80	20	87.56	71	80	20	86.57
21	80	20	88.06	72	80	20	87.06
22	80	20	88.06	73	80	20	87.06
23	80	20	88.06	74	80	20	87.06
24	80	20	88.06	75	80	20	87.06
25	80	20	88.06	76	80	20	87.06
26	80	20	87.56	77	80	20	87.06
27	80	20	87.56	78	80	20	87.06

k	Data Latih (%)	Data Uji (%)	Akurasi (%)	k	Data Latih (%)	Data Uji (%)	Akurasi (%)
28	80	20	87.56	79	80	20	87.06
29	80	20	87.56	80	80	20	87.06
30	80	20	87.06	81	80	20	87.06
31	80	20	87.06	82	80	20	87.06
32	80	20	87.06	83	80	20	87.06
33	80	20	87.06	84	80	20	87.06
34	80	20	87.06	85	80	20	87.06
35	80	20	87.06	86	80	20	87.06
36	80	20	87.06	87	80	20	87.06
37	80	20	87.06	88	80	20	87.06
38	80	20	87.06	89	80	20	87.06
39	80	20	87.06	90	80	20	87.56
40	80	20	87.06	91	80	20	87.06
41	80	20	87.06	92	80	20	87.56
42	80	20	87.06	93	80	20	87.56
43	80	20	87.06	94	80	20	87.56
44	80	20	87.06	95	80	20	87.56
45	80	20	87.06	96	80	20	87.56
46	80	20	87.06	97	80	20	87.56
47	80	20	86.57	98	80	20	87.56
48	80	20	86.57	99	80	20	87.56
49	80	20	86.57	100	80	20	87.56
50	80	20	86.57	<b>Akurasi rata-rata</b>			<b>87.16</b>
51	80	20	87.06	<b>Akurasi minimum</b>			<b>86.07</b>
52	80	20	87.06	<b>Akurasi maksimum</b>			<b>88.06</b>

Representasi Tabel 5.2 dalam bentuk gambar memudahkan pembaca dalam hal analisa hubungan nilai  $k$  dengan tingkat akurasi. Representasi tersebut disajikan ke dalam Gambar 5.1.

Dari Gambar 5.1 dapat dilihat bahwa akurasi maksimum yang didapat melalui jalur SBNMPTN/SNMPTN adalah 88,06% saat berada pada  $k = 7, 21, 22, 23, 24$  dan  $25$ , sedangkan akurasi minimumnya sebesar 86,07% saat berada pada  $k = 2$ . Gambar 5.1 juga memperlihatkan bahwa akurasi yang didapat cenderung stabil, kenaikan dan penurunannya tidak terlalu tajam.



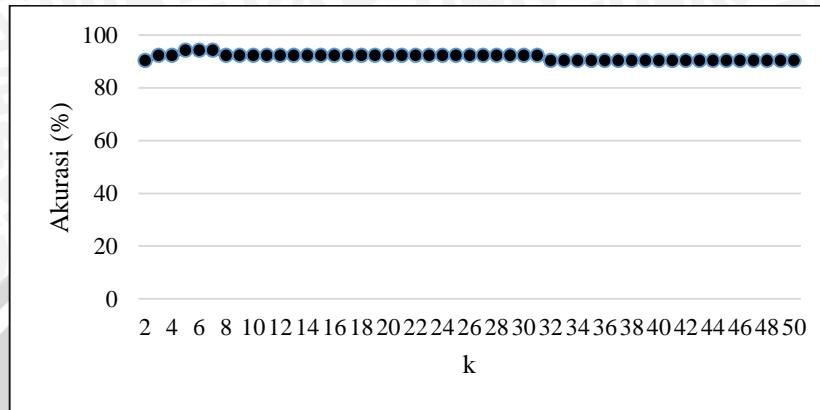


Gambar 5.1 Grafik Akurasi Skenario 1 Pada Jalur SNMPTN/SBNMPTN

Tabel 5.3 Hasil Akurasi Skenario 1 Pada Jalur SPMK

k	Data Latih (%)	Data Uji (%)	Akurasi (%)	k	Data Latih (%)	Data Uji (%)	Akurasi (%)
2	80	20	90.38	28	80	20	92.31
3	80	20	92.31	29	80	20	92.31
4	80	20	92.31	30	80	20	92.31
5	80	20	94.23	31	80	20	92.31
6	80	20	94.23	32	80	20	90.38
7	80	20	94.23	33	80	20	90.38
8	80	20	92.31	34	80	20	90.38
9	80	20	92.31	35	80	20	90.38
10	80	20	92.31	36	80	20	90.38
11	80	20	92.31	37	80	20	90.38
12	80	20	92.31	38	80	20	90.38
13	80	20	92.31	39	80	20	90.38
14	80	20	92.31	40	80	20	90.38
15	80	20	92.31	41	80	20	90.38
16	80	20	92.31	42	80	20	90.38
17	80	20	92.31	43	80	20	90.38
18	80	20	92.31	44	80	20	90.38
19	80	20	92.31	45	80	20	90.38
20	80	20	92.31	46	80	20	90.38
21	80	20	92.31	47	80	20	90.38
22	80	20	92.31	48	80	20	90.38
23	80	20	92.31	49	80	20	90.38
24	80	20	92.31	50	80	20	90.38
25	80	20	92.31	<b>Akurasi rata-rata</b>			<b>91.64</b>
26	80	20	92.31	<b>Akurasi minimum</b>			<b>90.38</b>
27	80	20	92.31	<b>Akurasi maksimum</b>			<b>94.23</b>

Representasi Tabel 5.3 dalam bentuk gambar memudahkan pembaca dalam hal analisa hubungan nilai  $k$  dengan tingkat akurasi. Representasi tersebut disajikan ke dalam Gambar 5.2.



Gambar 5.2 Grafik Akurasi Skenario 1 Pada Jalur SPMK

Dari Gambar 5.2 di atas dapat dilihat bahwa akurasi maksimum yang didapat melalui jalur SPMK adalah 94,23% saat berada pada  $k = 5, 6$  dan  $7$ , sedangkan akurasi minimumnya sebesar 90,38% saat berada pada  $k = 2$ . Gambar 5.2 juga memperlihatkan bahwa akurasi yang didapat cenderung stabil, kenaikan dan penurunannya tidak terlalu tajam.

## 2. Pengujian Skenario 2

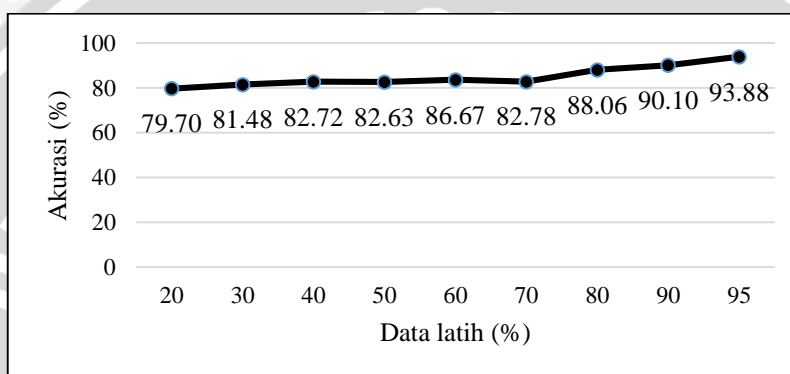
Pengujian skenario 2 melibatkan nilai  $k=6$  pada jalur SNMPTN/SBNMPTN dan  $k=6$  pada jalur SPMK. Pada skenario 2 nilai  $k$  konstan, sedangkan perbandingan data latih dan data ujinya berjalan. Hasil akurasi disajikan pada Tabel 5.4 berikut :

Tabel 5.4 Hasil Akurasi Skenario 2 Pada Jalur SBNMPTN/SNMPTN

<b>k</b>	<b>Data Latih (%)</b>	<b>Data Uji (%)</b>	<b>Akurasi (%)</b>
7	20	80	79.70
7	30	70	81.48
7	40	60	82.72
7	50	50	82.63
7	60	40	83.58
7	70	30	82.78
7	80	20	88.06
7	90	10	90.10

k	Data Latih (%)	Data Uji (%)	Akurasi (%)
7	95	5	93.88
<b>Akurasi rata-rata</b>			<b>84.99</b>
<b>Akurasi minimum</b>			<b>79.70</b>
<b>Akurasi maksimum</b>			<b>93.88</b>

Representasi Tabel 5.4 dalam bentuk gambar memudahkan pembaca dalam hal analisa hubungan nilai  $k$  dengan tingkat akurasinya. Representasi tersebut disajikan ke dalam Gambar 5.3.



Gambar 5.3 Grafik Akurasi Skenario 2 Pada Jalur SBNMPTN/SNMPTN

Pada Gambar 5.3 dapat dilihat bahwa akurasi maksimum yang bernilai 93,88% terjadi saat perbandingan data latih dan data uji sebesar 95% : 5% dari sejumlah 1.004 data. Sedangkan untuk akurasi minimumnya sebesar 79,70% terjadi saat perbandingan data latih dan data uji sebesar 20% : 80% dari sejumlah 1.004 data. Dari serangkaian pengujian tersebut di atas dapat disimpulkan bahwa semakin sedikit maupun semakin banyak data latihnya tidak menjamin bahwa akurasinya menurun ataupun meningkat. Jadi akurasi mampu mencapai optimal pada komposisi data yang sesuai. Untuk pengujian skenario 2 terhadap jalur SPMK tersaji dalam Tabel 5.5 berikut :

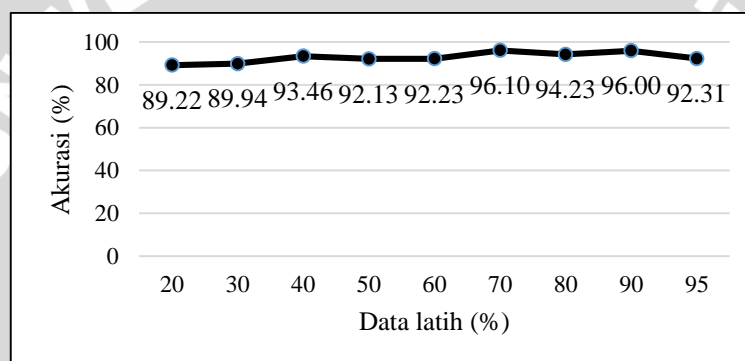
Tabel 5.5 Hasil Akurasi Skenario 2 Pada Jalur SPMK

k	Data Latih (%)	Data Uji (%)	Akurasi (%)
6	20	80	89.22
6	30	70	89.94
6	40	60	93.46
6	50	50	92.13
6	60	40	92.23



k	Data Latih (%)	Data Uji (%)	Akurasi (%)
6	70	30	96.10
6	80	20	94.23
6	90	10	96.00
6	95	5	92.31
<b>Akurasi rata-rata</b>			<b>92.85</b>
<b>Akurasi minimum</b>			<b>89.22</b>
<b>Akurasi maksimum</b>			<b>96.10</b>

Representasi Tabel 5.5 dalam bentuk gambar memudahkan pembaca dalam hal analisa hubungan nilai  $k$  dengan tingkat akurasi. Representasi tersebut disajikan ke dalam Gambar 5.4.



Gambar 5.4 Grafik Akurasi Skenario 2 Jalur SPMK

Hasil akurasi yang ada pada Gambar 5.4 menunjukkan bahwa akurasi maksimum dengan nilai 96,1% terjadi saat perbandingan data latih dan data uji sebesar 70% : 30% dari sejumlah 256 data. Sedangkan akurasi minimumnya dengan nilai 89,22% terjadi saat perbandingan data latih dan data ujinya sebesar 20% : 80% dari sejumlah 256 data. Hal ini dapat disimpulkan bahwa semakin sedikit dan semakin banyak data ujinya maka tidak menjamin akurasi menurun ataupun meningkat. Akurasi menurun saat prosentase data latihnya kurang dari 70% dari 256 data. Jadi akurasi mampu mencapai optimal saat berada pada komposisi data yang sesuai.

### 3. Pengujian Skenario 3

Pengujian pada skenario 3 merupakan pengujian keterlibatan fitur dalam perhitungan. Tujuan dari pengujian ini adalah untuk mencari akurasi maksimum dari perhitungan guna dipakai sebagai pertimbangan dalam menentukan kategori

uang kuliah baru yang dimasukkan oleh pengguna aplikasi. Pengujian fitur melibatkan 5 fitur utama saja yaitu penghasilan orang tua, biaya listrik, biaya air, biaya telepon dan Pajak Bumi dan Bangunan (PBB). Kemudian fitur tersebut direduksi satu persatu hingga kelimanya hilang (28 kali) untuk mengetahui pengaruh masing-masing fitur terhadap akurasi. Nilai  $k$  yang terlibat yaitu  $k=5$  untuk jalur SNMPTN/SBNMPTN dan  $k=3$  untuk jalur SPMK. Perbandingan data latih dan data uji untuk jalur SNMPTN/SBNMPTN data adalah 95% : 5% dari 1.004 data, sedangkan untuk jalur SPMK adalah 70% : 30% dari 256 data. Pengujiannya tersaji dalam Tabel 5.6 berikut.

Tabel 5.6 Pengujian Hasil Akurasi Skenario 3

Jumlah Fitur yang direduksi	Fitur yang Direduksi	Akurasi (%)		Jumlah Fitur yang direduksi	Fitur yang Direduksi	Akurasi (%)		
		SBNMPTN/	SPMK			SBNMPTN/	SPMK	
		SNMPTN				SNMPTN		
1	1	79.59	93.88	3	1,2,3	81.63	88.31	
	2	91.84	93.51		1,2,4	83.67	89.61	
	3	93.88	93.51		1,2,5	67.35	88.31	
	4	93.88	93.51		2,3,4	93.88	94.81	
	5	93.88	89.61		2,3,5	91.84	94.81	
2	1,2	75.51	88.31	4	2,4,5	93.88	94.81	
	1,3	85.71	93.51		3,4,5	93.88	94.81	
	1,4	79.59	92.21		1,2,3,4	81.63	89.61	
	1,5	75.51	92.21		1,2,3,5	57.14	87.01	
	2,3	91.84	94.81		1,2,4,5	63.27	87.01	
	2,4	91.84	94.81		1,3,4,5	81.63	90.91	
	2,5	93.88	92.21		2,3,4,5	91.84	94.81	
	3,4	91.84	96.1		5	1,2,3,4,5	51.02	87.01
	3,5	95.92	96.1		<b>Akurasi rata-rata</b>		<b>84.40</b>	<b>92.13</b>
	4,5	95.92	90.91		<b>Akurasi minimum</b>		<b>51.02</b>	<b>87.01</b>
					<b>Akurasi maksimum</b>		<b>95.92</b>	<b>96.10</b>

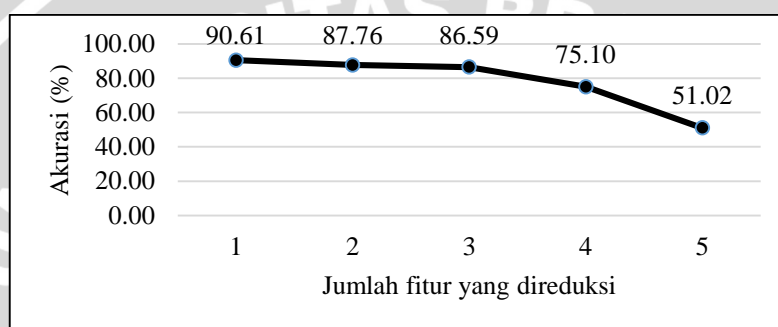
Keterangan :

1. Gaji orang tua
2. Biaya listrik
3. Biaya air
4. Biaya telepon
5. PBB

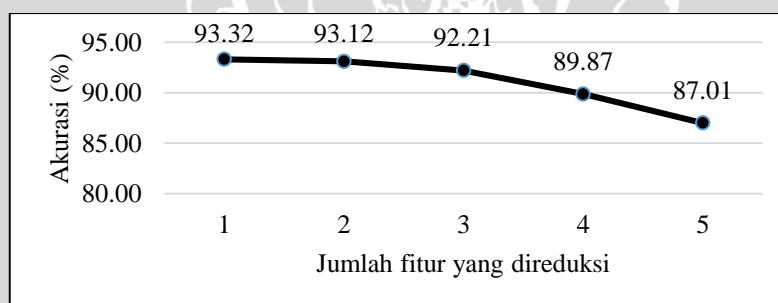


Representasi Tabel 5.6 dalam bentuk gambar memudahkan pembaca dalam hal analisa hubungan nilai  $k$  dengan tingkat akurasinya. Representasi tersebut disajikan ke dalam Gambar 5.5 dan Gambar 5.6.

Gambar 5.5 memuat hasil akurasi rata-rata dari pengujian skenario 3 di mana ada pereduksian fitur yang terlibat dalam pengklasifikasian. Rata-rata akurasi terbesar yang didapat dari jalur SNMPTN/SBNMPTN terjadi saat jumlah fitur yang direduksi hanya 1. Secara garis besar, semakin banyak fitur yang direduksi maka akurasinya akan semakin turun.



Gambar 5.5 Grafik Rata-rata Akurasi Skenario 3 Pada Jalur SBNMPTN/SNMPTN



Gambar 5.6 Grafik Rata-rata Akurasi Skenario 3 Pada Jalur SPMK

Gambar 5.6 memuat hasil akurasi rata-rata dari pengujian skenario 3 di mana ada pereduksian fitur yang terlibat dalam pengklasifikasian. Pada jalur SPMK didapat akurasinya yang semakin menurun dengan semakin banyaknya pereduksian fitur yang terjadi.

#### 4. Pengujian Skenario 4

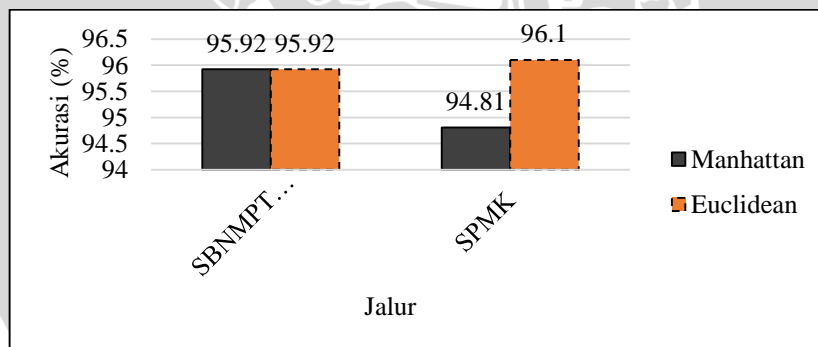
Pada pengujian ini melibatkan nilai  $k=7$  serta proporsi data latih dan data uji sebesar 95% : 5% dari 1.004 data pada jalur SBNMPTN/SNMPTN, sedangkan

$k=6$  serta proporsi data latih dan data uji sebesar 70% : 30% dari 256 data pada jalur SPMK. Hal yang diuji dari skenario 4 adalah metode perhitungan jaraknya. Jika pada skenario 1 hingga skenario 3 memakai metode jarak Euclidean maka pada skenario 4 menggunakan metode jarak Manhattan. Perhitungan jarak Manhattan dijabarkan seperti pada Persamaan 2.14 dengan  $p=1$ . Tujuan dari pengujian ini adalah untuk mencari akurasi maksimum dari perhitungan guna dipakai sebagai pertimbangan dalam menentukan kategori uang kuliah baru yang dimasukkan oleh pengguna aplikasi. Hasil perhitungan akurasi pada skenario 4 tersaji dalam Tabel 5.7 berikut :

Tabel 5.7 Hasil Pengujian Akurasi Skenario 4

Jalur	Perbandingan Jumlah Data		k	Akurasi (%)	
	Data latih (%)	Data uji (%)		Manhattan	Euclidean
SBNMPTN / SNMPTN	95	5	7	95,92	95,92
SPMK	70	30	6	94,81	96,1

Representasi Tabel 5.7 dalam bentuk gambar memudahkan pembaca dalam hal analisa hubungan nilai  $k$  dengan tingkat akurasinya. Representasi tersebut disajikan ke dalam Gambar 5.7.



Gambar 5.7 Grafik Akurasi Skenario 4

Gambar 5.7 memperlihatkan bahwa hasil akurasi yang didapat dengan menggunakan metode perhitungan jarak Manhattan memiliki hasil yang sama dengan jarak Euclidean untuk data melalui jalur SNMPTN dan SBNMPTN. Sedangkan untuk jalur SPMK, di antara kedua metode tersebut memiliki perbedaan akurasi. Akurasi terbesar diperoleh dengan menerapkan metode perhitungan jarak Euclidean.

Alasannya adalah bahwa jarak Euclidean merupakan metode perhitungan jarak yang benar-benar menghasilkan jarak terdekat, sehingga akurasi lebih bagus dibanding jarak Manhattan. Hal tersebut seperti diungkap dalam perhitungan Eugene Krause yang disunting ke dalam situs [www.taxicabgeometry.net](http://www.taxicabgeometry.net) bahwa perhitungan jarak pitagoras dengan menggunakan jarak Euclidean menghasilkan nilai yang lebih kecil (lebih dekat) dibanding dengan jarak Manhattan [TAX-14]. Menurut penjabaran Eko Prasetyo dalam bukunya yang berjudul *Data Mining* bahwa algoritma *Nearest Neighbor* melakukan klasifikasi berdasarkan kedekatan lokasi.

## 5.2 Analisa Hasil

Tabel 5.8 Hasil Akurasi Masing-masing kategori

Kategori	Total Data Uji	Jumlah		Akurasi (%)
		benar	salah	
1	17	14	3	0.82
2	4	0	4	0
3	2	0	2	0
4	14	8	6	0.57
5	10	4	6	0.4
6	154	151	3	0.98

Tabel 5.8 menunjukkan hasil akurasi yang didapat dari masing-masing kategori, melibatkan  $k=7$  dan jumlah data latih sebesar 80% dari 1.004 buah data dengan jumlah data yang tidak sama untuk masing-masing kategorinya. Tabel 5.8 tersebut menunjukkan bahwa hasil yang didapat dengan menerapkan *Fuzzy K-Nearest Neighbor* tidak disarankan bagi kasus yang memiliki jumlah data sedikit dengan sebaran yang tidak merata untuk masing-masing kategorinya, karena hal tersebut mengakibatkan hasil klasifikasi yang diperoleh akan cenderung menghasilkan kategori / kelas yang memiliki jumlah data paling banyak, dalam penelitian ini cenderung terklasifikasi ke dalam kelas 1, 4 dan 6.

Dari keseluruhan hasil akurasi yang terdapat beberapa faktor yang mampu mempengaruhi tingkat akurasi. Di antaranya adalah perbandingan jumlah data latih dan data uji yang terlibat, pemilihan nilai  $k$  dan pemakaian fitur.



Faktor pertama yaitu perbandingan jumlah data latih dan data uji, hasil pada Gambar 5.3 dan Gambar 5.4 menunjukkan bahwa akurasi yang didapat dari skenario 2 berubah-ubah setiap ada perubahan perbandingan jumlah data latih dan data ujinya. Begitu juga dengan  $k$ , nilai  $k$  mempengaruhi akurasi seperti terlihat pada Gambar 5.1 dan Gambar 5.2, pemakaian jumlah fitur mempengaruhi akurasi seperti terlihat pada Gambar 5.5 dan Gambar 5.6.

Akurasi maksimum yang didapat dari serangkaian pengujian sebesar 95,92% dengan  $k=7$  pada jalur SBNMPTN/SNMPTN. Sedangkan akurasi maksimum yang didapat melalui jalur SPMK sebesar 96,1% dengan  $k=6$ . Perolehan hasil perhitungan akurasi maksimum tersebut mampu dijadikan sebagai pertimbangan dalam perancangan aplikasi akhir yang dipakai langsung oleh para pengguna sistem. Komposisinya berupa nilai  $k=7$  untuk jalur SBNMPTN/SNMPTN, dilakukan pada perbandingan data latih dan data uji sebesar 95% : 5% dari sejumlah 1.004 data dengan menghilangkan fitur tagihan telepon dan besar PBB, serta nilai  $k=6$  untuk jalur SPMK dan dilakukan pada perbandingan data latih dan data ujinya sebesar 70% : 30% dari sejumlah 256 data.