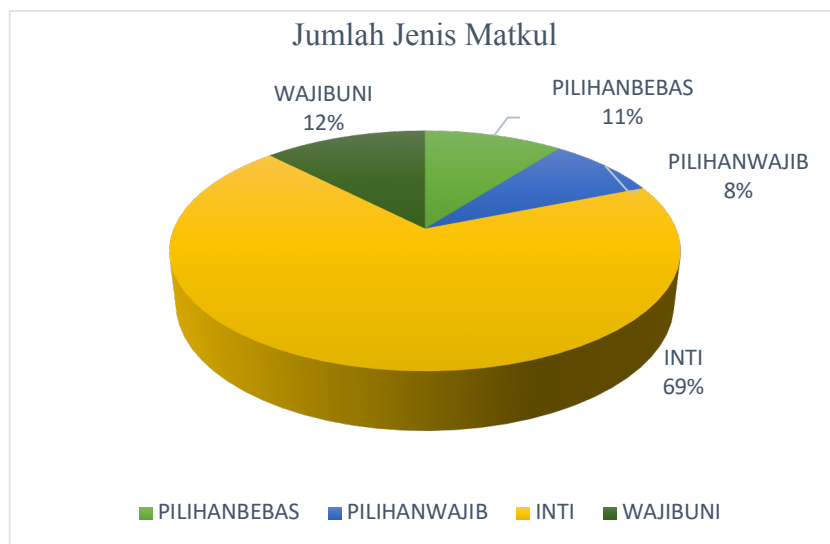


BAB V HASIL DAN PEMBAHASAN

Pada bab ini akan dijelaskan terkait hasil dan pembahasan pada klasifikasi nilai mata kuliah mahasiswa, terlebih dahulu maka dilakukan statistika deskriptif guna memberikan gambaran umum keadaan dataset mahasiswa yang digunakan dalam klasifikasi, berikut adalah deskriptifnya:

5.1. Statistika Deskriptif

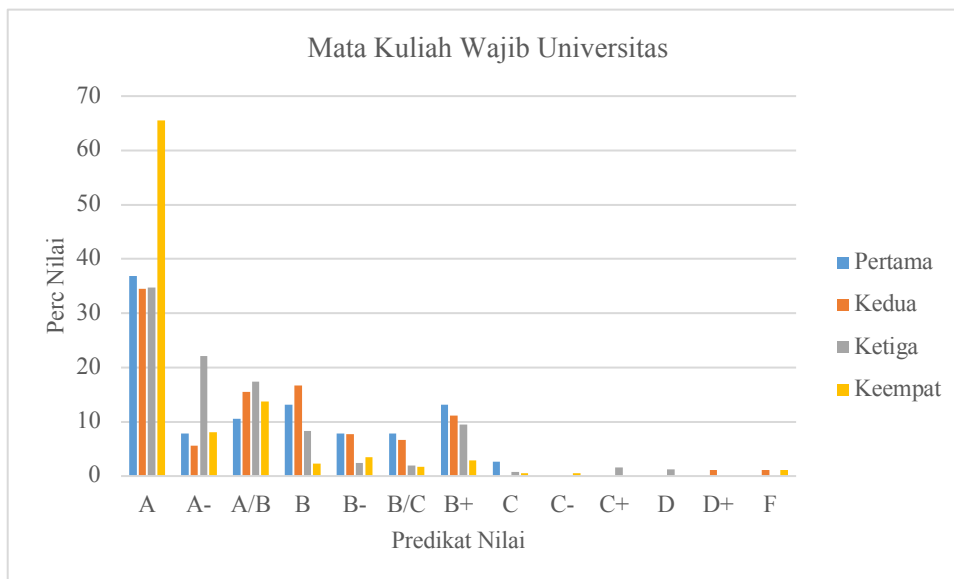
Berikut adalah deskripsi dengan bentuk *pie chart* mengenai jumlah total mata kuliah yang diambil oleh mahasiswa statistika angkatan 2015 dari tahun ajaran Genap 2015/ 2016 sampai Genap 2017/ 2018, yaitu:



Gambar 5.1 Diagram Jenis Mata Kuliah

Dari gambar 5.1 yang dirujuk menunjukkan bahwa jenis mata kuliah yang paling banyak diambil oleh mahasiswa statistika adalah 69% mata kuliah inti, 12% mata kuliah wajib universitas, 11% mata kuliah pilihan bebas, dan 8% mata kuliah pilihan wajib.

Kemudian berikut adalah diagram batang mata kuliah wajib universitas dengan jumlah perolehan masing – masing nilai mata kuliah berdasarkan jadwal jam pelajaran yang diambil mahasiswa, yaitu:

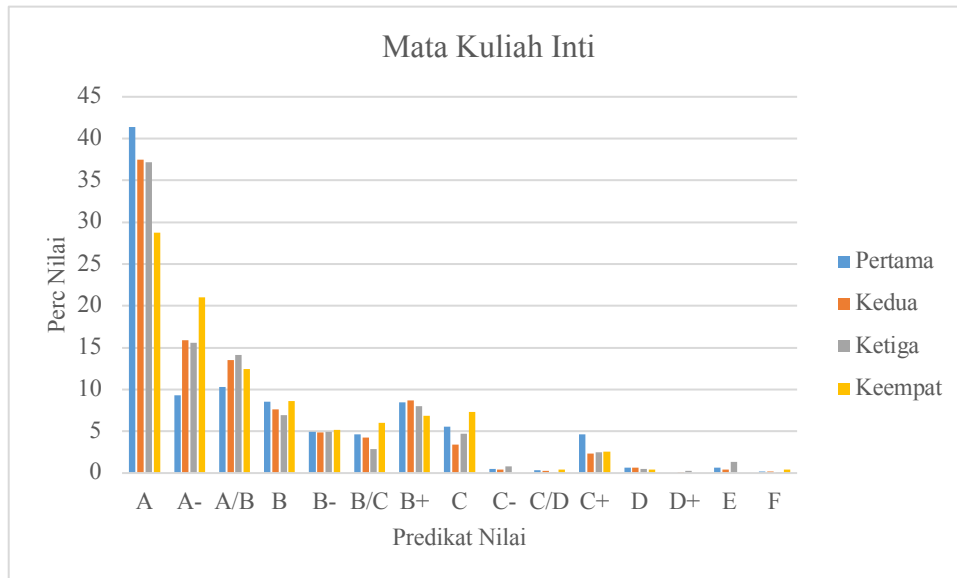


Gambar 5.2 Diagram Jumlah Mata Kuliah Wajib terhadap Nilai

Pada gambar 5.2 merupakan diagram batang dari seberapa persen mahasiswa memperoleh nilai pada saat mengambil jenis mata kuliah wajib universitas, dapat diketahui bahwa mahasiswa yang mendapatkan nilai A terbanyak yaitu mahasiswa yang mengambil jenis mata kuliah ini pada jam pelajaran keempat, dan terbanyak kedua yaitu mahasiswa yang mengambil pada jam kuliah pertama, terbanyak ketiga yaitu yang mengambil pada jam ketiga dan terakhir keempat.

Karena mata kuliah wajib Universitas ini juga banyak yang memiliki jumlah 2 SKS dengan materi kuliah yang tidak terlalu berat maka sangat cocok apabila diletakkan dalam jam kuliah keempat dan disusul dengan jam pertama, sehingga banyak mahasiswa yang mendapat nilai A jika masuk kedalam kelas jam keempat dan pertama.

Berikut adalah diagram batang mata kuliah inti dengan jumlah perolehan masing – masing nilai mata kuliah berdasarkan jadwal jam pelajaran yang diambil mahasiswa, yaitu:

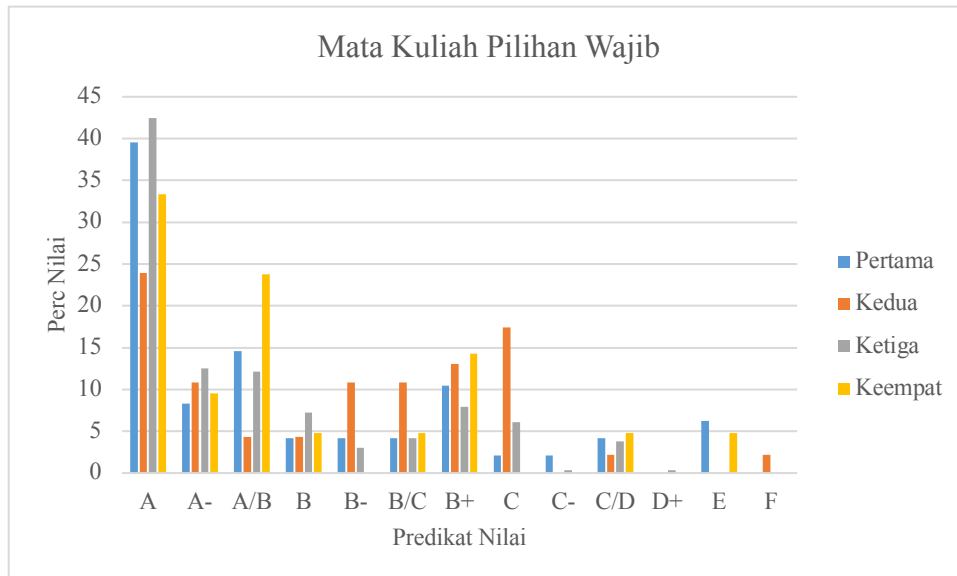


Gambar 5.3 Diagram Jumlah Matkul Inti terhadap Nilai

Pada gambar 5.3 merupakan diagram batang dari seberapa persen mahasiswa memperoleh nilai pada saat mengambil jenis mata kuliah inti, dapat diketahui bahwa mahasiswa yang mendapatkan nilai A terbanyak yaitu mahasiswa yang mengambil jenis mata kuliah inti pada kelas dengan jam pelajaran pertama, dan terbanyak kedua yaitu mahasiswa yang mengambil kelas pada jam kuliah kedua, terbanyak ketiga yaitu yang mengambil kelas pada jam ketiga dan terakhir keempat.

Hal ini sesuai dengan pengalaman peneliti bahwa dikarenakan mata kuliah inti Statistika merupakan mata kuliah yang membutuhkan konsentrasi yang tinggi maka sangat cocok diletakkan pada jam pagi yaitu jam pertama dan kedua, hal tersebut sangatlah efektif sehingga menjadikan kelas jam pertama dan kedua banyak mahasiswa yang memperoleh nilai A.

Kemudian berikut adalah diagram batang mata kuliah pilihan wajib dengan jumlah perolehan masing – masing nilai mata kuliah berdasarkan jadwal jam pelajaran yang diambil mahasiswa, yaitu:

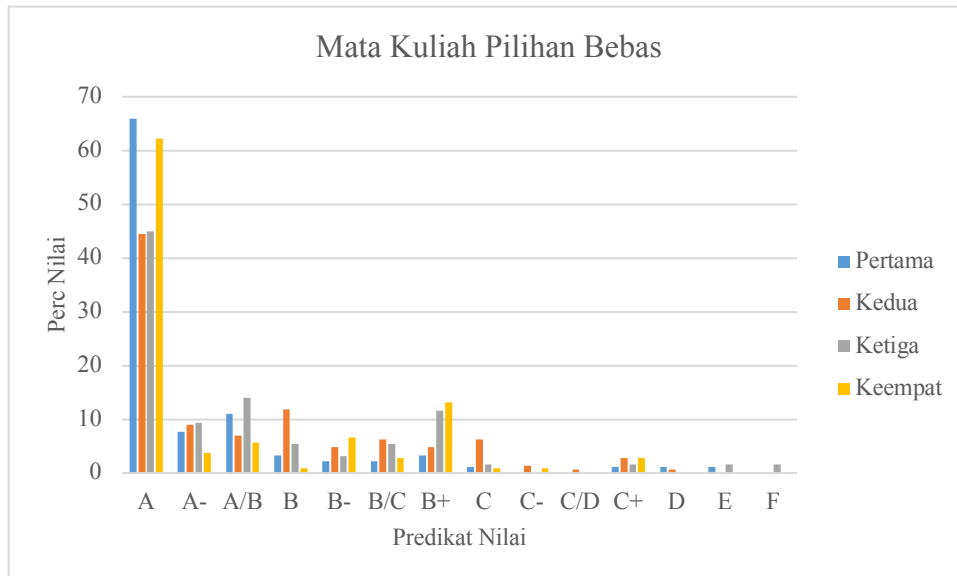


Gambar 5.4 Diagram Jumlah Mata Kuliah Pilihan Wajib terhadap Nilai

Pada gambar 5.4 merupakan diagram batang dari seberapa persen mahasiswa memperoleh nilai pada saat mengambil jenis mata kuliah pilihan wajib, dapat diketahui bahwa mahasiswa yang mendapatkan nilai A terbanyak yaitu mahasiswa yang mengambil jenis mata kuliah pilihan wajib pada kelas dengan jam pelajaran ketiga, dan terbanyak kedua yaitu mahasiswa yang mengambil kelas pada jam kuliah pertama, terbanyak ketiga yaitu yang mengambil kelas pada jam keempat dan terakhir kedua.

Berdasarkan pengalaman peneliti bahwa pada kelas jam ketiga, banyak mahasiswa dengan kelas tersebut mendapat nilai A karena pelajaran matkul pilihan wajib biasanya tidak terlalu berat, sehingga mahasiswa lebih memilih kelas dijam tersebut dan juga mahasiswa meletakkan slot jam pertama atau kedua untuk mata kuliah inti.

Kemudian berikut adalah diagram batang mata kuliah pilihan bebas dengan jumlah perolehan masing – masing nilai mata kuliah berdasarkan jadwal jam pelajaran yang diambil mahasiswa, yaitu:



Gambar 5.5 Diagram Jumlah Mata Kuliah Pilihan Bebas terhadap Nilai

Pada gambar 5.5 merupakan diagram batang dari seberapa persen mahasiswa memperoleh nilai pada saat mengambil jenis mata kuliah pilihan bebas, dapat diketahui bahwa mahasiswa yang mendapatkan nilai A terbanyak yaitu mahasiswa yang mengambil jenis mata kuliah pilihan bebas pada kelas dengan jam pelajaran pertama, dan terbanyak kedua yaitu mahasiswa yang mengambil kelas pada jam kuliah keempat, terbanyak ketiga yaitu yang mengambil kelas pada jam ketiga dan terakhir kedua.

Berdasarkan pengalaman peneliti, untuk jenis mata kuliah bebas mahasiswa memprioritaskan untuk mengambil kelas jam keempat dan jika kehabisan kelas maka pilihan yang lain yaitu mahasiswa banyak mengambil kelas pada jam pertama. Hal tersebut dilakukan karena mata kuliah pilihan bebas memiliki materi yang cukup ringan dan sesuai dengan konsentrasi yang diambil masing – masing mahasiswa, sehingga pada kedua jam tersebut mahasiswa banyak mendapat nilai A.

Kemudian berikut adalah pembahasan mengenai klasifikasi nilai mata kuliah mahasiswa dengan menggunakan SVM:

5.2. Klasifikasi *Support Vector Machine*

Didalam penelitian ini *grade* nilai mahasiswa Statistika angkatan 2015 akan diklasifikasi menjadi 6 kelas secara berturut – turut dengan diberi nama variabel

predikat sesuai dengan klasifikasi di tabel 4.1 , variabel predikat ini sebagai variabel dependen. Sedangkan variabel independen yg digunakan pada penelitian ini adalah jenis mata kuliah, jam mata kuliah, hari, jumlah SKS sudah dikategorikan sesuai dengan data, serta variabel numerik yaitu presentase keterlambatan mahasiswa.

Sebelum melakukan klasifikasi terlebih dahulu menentukan data *training* dan data *testing* untuk memprediksikan kelas. Dalam penelitian ini sebagian dari jumlah data set diambil untuk data *training* sebanyak 66,67% dari total data dan diambil secara random, karena pengambilan data *training* yaitu dengan cara random maka dilakukanlah *running set seed* terlebih dahulu pada saat sebelum merandom data agar pengambilan random hanya dilakukan satu kali saja.

Setelah ditentukan data *training* maka selanjutnya menentukan data *testing*. Dalam penelitian ini sisa dari data *training* yaitu 33,33% akan digunakan dalam data *testing*. Data *testing* digunakan untuk mengukur sejauh mana *classifier* berhasil melakukan klasifikasi dengan benar. Karena itu, data yang ada pada data *testing* seharusnya tidak boleh ada pada *training* set sehingga dapat diketahui apakah model *classifier* sudah “pintar” dalam melakukan klasifikasi.

Memisahkan data menjadi *training* dan *testing* dimaksudkan agar model yang diperoleh nantinya memiliki kemampuan generalisasi yang baik dalam melakukan klasifikasi data. Tidak jarang sebuah model klasifikasi dapat melakukan klasifikasi data dengan sangat baik pada data *training*, tetapi sangat buruk dalam melakukan klasifikasi data yang baru dan belum pernah ada. Hal ini dinamakan *overfitting*. Di dalam penelitian ini jumlah sampel yang digunakan yaitu 1497 dan data sampel diambil secara random, sampel tersebut juga akan digunakan sebagai data *testing*. Sedangkan sisanya digunakan sebagai data *training* sebanyak 2995 data.

Setelah dilakukan pembagian data *training* dan *testing*, maka dilakukan klasifikasi dengan dan tanpa metode SMOTE, yaitu sebagai berikut:

5.2.1 Menggunakan SMOTE

Setelah dilakukan pembagian data *training* dan data *testing*, maka dilakukan klasifikasi.

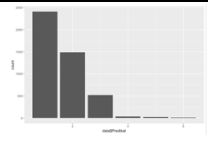
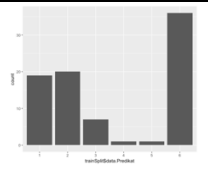
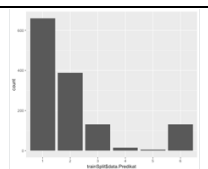
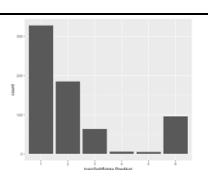
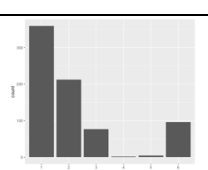
Pada penelitian ini terdapat perbedaan kelas yang sangat signifikan oleh karena itu maka dilakukan resampling data dengan metode SMOTE. Berikut merupakan hasil dari metode SMOTE. Metode ini bekerja dengan cara jika suatu variabel

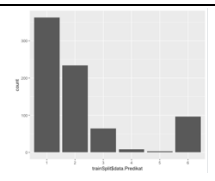
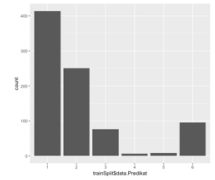
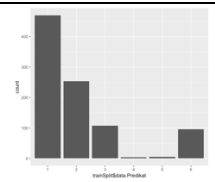
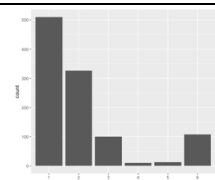
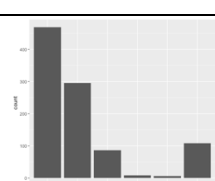
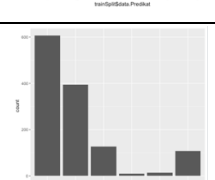
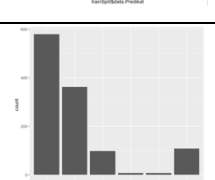
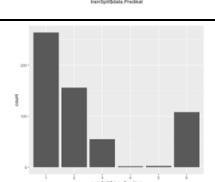
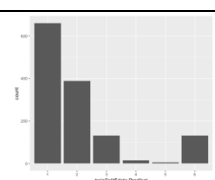
memiliki jumlah frekuensi yang banyak maka jumlah data akan dikurangi sebanyak 2% data dengan fungsi “*prec under*” dan sebaliknya jika jumlah suatu variabel paling sedikit maka data frekuensi akan ditambah sebanyak 2% data dengan fungsi “*prec over*”.

Setelah SMOTE selesai maka langsung dilanjutkan dengan klasifikasi SVM dengan hasil akurasi yaitu 44% hasil tersebut menunjukkan bahwa hasil klasifikasi tidak terlalu bagus karena tidak mendekati 100%. Hasil akurasi yang didapat sedikit dikarenakan terlalu banyak kelas dalam suatu variabel.

Karena hasil akurasi yang sedikit dan jumlah total data training yang menjadi sangat sedikit dari data awal, maka untuk menentukan jumlah *perc over* dan *perc under* yang cocok maka dilakukanlah beberapa percobaan yaitu sebagai berikut:

Tabel 5.1 Perbandingan *Perc Over* dan *Per Under* pada SMOTE

Perc Over	Perc Under	akurasi	Total data	Plot
Plot sebelum melakukan SMOTE				
2%	2%	0.44	63	
10%	10%	0.29	189	
7.5%	7%	0.56	513	
7.5%	7.8%	0.57	564	

Perc Over	Perc Under	akurasi	Total data	Plot
7.5%	8%	0.57	576	
7.5%	9%	0.59	639	
7.5%	10%	0.56	702	
8%	10%	0.59	801	
8%	9%	0.57	729	
8%	12%	0.57	945	
8%	11%	0.58	873	
8%	5%	0.56	441	
8.3%	5%	0.51	441	

Dari tabel 5.1 bahwa dapat disimpulkan bahwa range antara *perc over* dan *under* tidak boleh terlalu besar, *range* yang paling baik yaitu 2%. Selain dengan membandingkan akurasi dan jumlah total data juga membandingkan distribusi hasil sampling, maka peneliti memilih menggunakan *perc over* 8% dan *perc under* 10% dengan tingkat akurasi 0.59 dan total data train menjadi 801.

Hasil akurasi tersebut menggunakan *Cost* yaitu 10 dan *gamma* yaitu 1. Setelah selesai menentukan *perc over* dan *perc under* maka selanjutnya melakukan *running* fungsi SVM dan melihat modelnya dengan menentukan nilai *Cost* dan *Gamma* terbaik, yaitu sebagai berikut:

Dalam menentukan permodelan SVM maka terlebih dahulu tentukan nilai C atau *Cost* dan menentukan nilai *Gamma*.

Nilai C yang biasa dipakai dalam penelitian yaitu antara 0.01, 0.1, 1, 100, serta nilai *Gamma* yang biasa dipakai dalam penelitian yaitu antara 0.01, 0.1, 1, 100. Sehingga dalam penelitian ini akan dilakukan kombinasi dari nilai masing – masing parameter C dan *gamma* sebanyak 20 percobaan.

Maka peneliti menggunakan kombinasi dari *cost* dan *gamma* dan mencobanya pada data, dengan perbandingan yaitu sebagai berikut:

Tabel 5.2 Perbandingan *Cost* dan *Gamma* pada SMOTE

Cost	Gamma	Akurasi	Error Rate
0.01	0.01	0.49	0.51
	0.1	0.49	0.51
	1	0.49	0.51
	100	0.49	0.51
0.1	0.01	0.494	0.506
	0.1	0.477	0.523
	1	0.477	0.523

	100	0.477	0.523
100	0.01	0.493	0.507
	0.1	0.554	0.446
	1	0.575	0.425
	100	0.580	0.42
1	0.01	0.494	0.506
	0.1	0.509	0.491
	1	0.563	0.437
	100	0.578	0.422
10	0.01	0.480	0.52
	0.1	0.530	0.47
	1	0.580	0.42
	100	0.580	0.42

Karena pada saat *running* terdapat random maka dapat terjadi perbedaan akurasi akan tetapi hasil beberapa *running* akan menunjukkan hasil yang tidak jauh berbeda. Pada tabel 5.2 terdapat tiga nilai akurasi tertinggi yaitu 0.58 yaitu dengan *cost* dan gamma masing – masing 100 dan 100, 10 dan 1, serta 10 dan 100. Sehingga perlu dilakukan *running confusion matrix* untuk melihat secara detail hasil klasifikasinya yaitu sebagai berikut:

Tabel 5.3 *Confusion Matrix* SMOTE cost 10 dan gamma 1

Prediksi	Amat Baik	Baik	Cukup	Kurang	Sangat Kurang	Tdk Memenuhi
Amat Baik	353	93	45	2	2	11

Baik	39	112	23	2	4	4
Cukup	4	10	29	0	0	1
Kurang	0	0	0	0	0	0
Sangat Kurang	0	0	0	0	0	0
Tdk Memenuhi	1	1	0	0	0	65

Tabel 5.4 *Confusion Matrix* SMOTE cost 10 & gamma100 dan cost 100 & gamma 100

Prediksi	Amat Baik	Baik	Cukup	Kurang	Sangat Kurang	Tdk Memenuhi
Amat Baik	363	117	50	5	3	11
Baik	30	102	12	2	0	3
Cukup	2	4	26	0	0	1
Kurang	0	0	0	2	0	0
Sangat Kurang	0	0	0	0	1	0
Tdk Memenuhi	0	0	0	0	0	66

Berdasarkan perbandingan pada tabel 5.3 dengan maka *cost* 10 dan *gamma* 1, tidak ada data yang masuk kedalam kelas “kurang” dan “sangat kurang”, maka hasil klasifikasi tidak baik. Sementara itu hasil *confusion matrix* tabel 5.4 pada *cost* 100 *gamma* 100 dan *cost* 10 *gamma* 100 memiliki semua anggota pada setiap kelas dan memiliki nilai yang sama maka peneliti mengambil salah satu nilai *cost* dan *gamma* yaitu mahasiswa untuk *cost* yaitu 10 dan *gamma* yaitu 100.

Setelah klasifikasi SVM menggunakan metode SMOTE. Selanjutnya peneliti akan melakukan klasifikasi SVM tanpa menggunakan SMOTE, hal tersebut dilakukan untuk membandingkan hasil klasifikasi SVM dengan atau tanpa SMOTE, yaitu sebagai berikut:

5.2.2. Tanpa Menggunakan SMOTE

Berikut adalah pembahasan mengenai klasifikasi SVM nilai mata kuliah tanpa menggunakan SMOTE. Ada sedikit perbedaan pada data set saat di-*input*-kan kedalam *software* R yaitu variabel data Predikat sebelumnya jika menggunakan SMOTE maka variabel Predikat diubah terlebih dahulu dari kategorik menjadi numerik, akan tetapi jika tidak menggunakan SMOTE maka variabel Predikat tetap dengan variabel kategorik berupa amat baik, baik, cukup, kurang, sangat kurang dan tidak memenuhi.

Klasifikasi SVM tanpa menggunakan SMOTE sama seperti sebelumnya dengan menggunakan SMOTE pada pembagian data *training* dan *testing* menggunakan Teknik yang sama yaitu dengan mengambil sebagian dari jumlah data set untuk data *training* sebanyak 66,67% dari total data dan diambil secara random, karena pengambilan data *training* yaitu dengan cara random maka dilakukanlah *running set seed* terlebih dahulu pada saat sebelum merandom data agar pengambilan random hanya dilakukan satu kali saja. Setelah ditentukan data *training* maka selanjutnya menentukan data *testing*. Dalam penelitian ini sisa dari data *training* yaitu 33,33% akan digunakan dalam data *testing*.

Setelah ditentukan data *training* dan *testing* maka selanjutnya dilakukan permodelan SVM. Dalam menentukan permodelan SVM maka terlebih dahulu menentukan nilai C atau Cost dan menentukan nilai Gamma. Sama seperti sebelumnya, maka peneliti menggunakan kombinasi Cost dan Gamma untuk mencari klasifikasi yang terbaik, yaitu sebagai berikut:

Tabel 5.5 Perbandingan *Cost* dan *Gamma* tanpa SMOTE

Cost	Gamma	Akurasi	Error Rate
0.01	0.01	0.535	0.465
	0.1	0.535	0.465

Cost	Gamma	Akurasi	Error Rate
	1	0.535	0.465
	100	0.535	0.465
0.1	0.01	0.535	0.465
	0.1	0.535	0.465
	1	0.554	0.446
	100	0.554	0.446
100	0.01	0.559	0.441
	0.1	0.577	0.423
	1	0.590	0.41
	100	0.595	0.405
1	0.01	0.535	0.465
	0.1	0.558	0.442
	1	0.579	0.421
	100	0.596	0.404
10	0.01	0.535	0.465
	0.1	0.568	0.432
	1	0.584	0.416
	100	0.595	0.405

Karena pada saat *runing* pengambilan sampel dilakukan secara random sehingga dapat terjadi perbedaan akurasi akan tetapi hasil beberapa *running* akan menunjukkan nilai yang tidak signifikan berbeda. Berdasarkan perbandingan pada tabel dirujuk maka *cost* dan *gamma* terbaik yang digunakan untuk klasifikasi SVM data set nilai mahasiswa tanpa menggunakan SMOTE untuk *cost* yaitu 1 dan *gamma* yaitu 100 dengan akurasi 0.596 dengan *confussion matriks* yaitu sebagai berikut:

Tabel 5.6 *Confusion Matrix* tanpa SMOTE

Prediksi	Amat Baik	Baik	Cukup	Kurang	Sangat Kurang	Tdk Memenuhi
Amat Baik	1419	603	201	17	14	5
Baik	177	368	120	1	8	2
Cukup	9	10	37	1	0	0
Kurang	0	0	0	0	0	0
Sangat Kurang	0	0	0	2	0	0
Tdk Memenuhi	0	0	0	0	0	2

Setelah dilakukan klasifikasi SVM dengan dan tanpa menggunakan SMOTE maka selanjutnya membandingkan hasil dari keduanya, yaitu sebagai berikut:

5.3. Perbandingan Dua Metode

Setelah melakukan klasifikasi SVM dengan dan tanpa menggunakan metode SMOTE, maka selanjutnya membandingkan hasil dari kedua metode tersebut didalam tabel 5.7 dibawah:

Tabel 5.7 Perbandingan SMOTE dan Tanpa SMOTE

	SMOTE	TANPA SMOTE
<i>Accuracy</i>	0.58	0.59
<i>Error rate</i>	0.42	0.41
<i>Cost</i>	10	1
<i>Gamma</i>	100	100

Jika dilihat dari hasil akurasi dan *error* maka kedua metode sama – sama baik dan tidak jauh berbeda, akan tetapi jika dilihat dari tidak seimbangnya data set pada plot sebelum menggunakan SMOTE pada tabel 5.1 maka peneliti lebih memilih klasifikasi SVM dengan menggunakan metode SMOTE.

Selain menggunakan *accuracy* dan *error* maka dilihat juga *precision* dan *recall*-nya pada tabel 5.8 yaitu sebagai berikut:

Tabel 5.8 *Precision* dan *Recall*

Variabel kelas	Dengan SMOTE		Tanpa SMOTE	
	<i>Precision</i>	<i>Recall</i>	<i>Precision</i>	<i>Recall</i>
Amat baik	0.661	0.918	0.628	0.884
Baik	0.684	0.457	0.544	0.375
Cukup	0.787	0.295	0.649	0.103
Kurang	1	0.222	-	-
Sangat kurang	1	0.25	0	0
Tidak memenuhi	1	0.814	1	0.222

Dilihat dari tabel 5.8 pada *confusion matrix* tanpa menggunakan SMOTE kelas dengan nilai predikat “Kurang” tidak ada, dan hasil *presicion* dan *recall* dari *confusion matrix* dengan menggunakan SMOTE lebih bagus dari pada tanpa menggunakan SMOTE.