

Segmentasi Dan Klasifikasi Perilaku Pembayaran Pelanggan Pada Perusahaan Multimedia Dengan Algoritma K-Means Dan C4.5

Ela Nurelasari

Program Studi Sistem Informasi Universitas Bina Sarana Informatika
e-mail: ela.eur@bsi.ac.id

Cara Sitasi: Nurelasari, E. (2019, Maret). Segmentasi Dan Klasifikasi Perilaku Pembayaran Pelanggan Pada Perusahaan Multimedia Dengan Algoritma K-Means Dan C4.5. (S. Dalis, Ed.) Paradigma - Jurnal Komputer dan Informatika, 21(1), 69-76. doi:10.31294/p.v21i1.5042

Abstract - A service provider company often have problem with insolvent customer. As a company that can not verify customer financial, a service provider company is having difficulty on identifying solvent customer. As one of the service provider company in Jakarta which offer internet connectivity and cable TV service, identifying solvent customer is important because, every customer inside coverage area is a potential customer to subscribe their product. In a related research, algorithm C4.5 become one of the favorite algorithm to use in classifying customer insolvency, and was able to produce high accuracy model, and easy to understand. The major problem of algorithm C4.5 is attribute selection. As noted that a service provider company cannot verify customer financial, they have no data about customer financial which is one important attribute in credit scoring. Therefore, customer payment behavior will be extracted as customer potential as one of classification attribute. Feature extraction with k-means was able to improve the accuracy of the algorithm C4.5. As noted by improve on accuracy from 59.02% to 77.31% and AUC from 0.537 to 0.836. The customer potential attribute can also be used as reference in the promotion, retention, and managing insolvent customer.

Keywords: Customer insolvency, customer potential, algorithm C4.5, algorithm K-Means

PENDAHULUAN

Salah satu perusahaan pelayanan jasa yang bergerak dalam layanan tv kabel dan internet di Jakarta. Memberikan layanan kepada pelanggan dengan infrastruktur kabel *fiber optic* ke areal perumahan. Pelanggan yang masuk kedalam jaringan pelayanan, akan dihubungi dan ditawarkan promosi penggunaan jasa. Penawaran diberikan kepada setiap pelanggan yang masuk dalam jangkauan tanpa verifikasi detail terlebih dahulu. Jika pelanggan yang sudah berlangganan terlambat atau menolak membayar, perusahaan tetap memberikan jasa dengan pembatasan tertentu selama 3 bulan (pembatasan *channel* untuk tv kabel, dan pembatasan kecepatan untuk internet) dengan harapan pelanggan dapat terus menggunakan jasa.

Perusahaan penyedia layanan multimedia sering mengalami masalah dengan pelanggan yang menolak melakukan pembayaran atas jasa yang sudah digunakan (Daskalaki, Kopanas, Goudara, & Avouris, 2003). Berbeda dengan bank atau perusahaan yang menawarkan pinjaman dana, perusahaan jasa yang pascabayar seringkali memberikan jasanya kepada pelanggan tanpa melalui proses verifikasi yang detail, sehingga sulit

untuk mengetahui apakah pelanggan yang berlangganan merupakan pelanggan yang layak (Pinheiro, Evsukoff, & Ebecken, 2006). Oleh karena itu terdapat pelanggan yang menolak melakukan pembayaran sehingga menimbulkan hutang dan mengurangi keuntungan.

Agar menekan resiko kurang tepatan dalam memprediksi resiko maka diperlukan suatu cara yang baik untuk menangani hal tersebut (Sari & Sindunata, 2014). Perusahaan jasa memiliki regulasi untuk tetap melanjutkan layanan dalam periode tertentu walaupun pelanggan menolak membayar (Pinheiro et al., 2006). Meskipun adanya penalti yang dikenakan, namun tetap menjadi masalah. Deteksi dan pencegahan perilaku pelanggan yang menolak membayar ini adalah salah satu tujuan yang penting untuk dicapai dalam industri.

Dalam klasifikasi kelayakan, salah satu atribut yang sangat berpengaruh adalah keuangan pelanggan. Namun perusahaan penyedia layanan multimedia tidak memiliki data detail mengenai keuangan pelanggan (Daskalaki et al., 2003). Oleh karena itu data pembayaran pelanggan dapat disegmentasi untuk melihat potensial pelanggan dan membantu perusahaan dalam melakukan tindakan pencegahan yang tepat sesuai dengan

segmen pelanggan (Pinheiro et al., 2006). Dengan demikian, perusahaan dapat mengambil tindakan yang sesuai dengan kelompok pelanggan.

Hubungan antara pelanggan dengan perusahaan sangat penting untuk menunjang perkembangan dan kelangsungan perusahaan, pihak manajemen perusahaan harus mampu untuk mengenali pelanggan potensialnya dan mempercayainya dengan meningkatkan pemahaman perusahaan akan kebutuhan mereka sebagai individu sehingga dapat mempertahankan loyalitasnya terhadap perusahaan (Putra & Wadisman, 2018).

Metode *data mining* saat ini sudah banyak digunakan untuk meningkatkan bisnis dan meningkatkan keuntungan dalam kompetisi dengan perusahaan pesaing. *Data mining* dapat mengolah data dalam jumlah besar dan menghasilkan pengetahuan yang dijadikan acuan dalam keputusan bisnis (Thiruvadi & Patel, n.d.). *Data mining* dapat digunakan untuk klasifikasi dan segmentasi. Teknik segmentasi merupakan teknik yang memecah data kedalam kelompok-kelompok yang homogen. Sedangkan teknik klasifikasi bertujuan untuk melihat perilaku pelanggan yang tidak layak terutama saat menjelang periode pembayaran.

Data mining sudah banyak digunakan untuk masalah perilaku pelanggan, (Liao & Triantaphyllou, n.d.) menunjukkan banyak penelitian mengenai data mining, dimana penelitian yang melibatkan pelanggan menjadi salah satu kategori besar. (Thiruvadi & Patel, n.d.) melakukan survei terhadap penerapan *data mining* dalam deteksi dan pencegahan kecurangan dimana disebutkan bahwa pelanggan yang memakai jasa dan menolak membayar juga tergolong dalam kecurangan.

Pelanggan perusahaan penyedia jasa yang terlambat atau menolak membayar memberikan dampak negatif bagi perusahaan (Daskalaki, Kopanas, (Daskalaki et al., 2003), oleh karena itu banyak penelitian yang menggunakan *machine learning* untuk memprediksi pelanggan yang tidak baik, serta melakukan tindakan pencegahan. C4.5 merupakan salah satu algoritma yang populer dengan tingkat akurasi yang tinggi pada penelitian serupa. Namun jika digunakan untuk data dengan atribut yang banyak, serta kurang relevan dan redundan dapat mengurangi akurasi algoritma C4.5 (Liu & Schumann, n.d.). Oleh karena itu perlu dilakukan pemilihan atau pengelompokan atribut untuk mengurangi atribut yang redundan.

Perilaku pelanggan perlu dipelajari agar kebijakan yang lebih sesuai dapat diambil, oleh karena itu perilaku pelanggan akan disegmentasi dengan K-Means. Selain itu segmentasi perilaku ini juga membantu mengurangi atribut pembayaran yang cenderung redundan. Pemrosesan data awal dengan *k-means* dapat meningkatkan akurasi dari algoritma klasifikasi (Ali, Sulaiman, Mustapha, & Mustapha,

2009). Selain itu dengan mempelajari perilaku pelanggan, kebijakan yang dibuat dapat lebih spesifik untuk mempertahankan serta meningkatkan relasi dengan pelanggan.

(Daskalaki et al., 2003) menerapkan data mining untuk membantu memprediksi pelanggan yang menolak membayar pada perusahaan telekomunikasi. (Pinheiro et al., 2006) menggunakan *self organizing map* untuk segmentasi pelanggan dan diprediksi dengan *neural network* untuk menentukan kelayakan pelanggan. (Prasad, 2011) melakukan segmentasi pelanggan untuk mendapatkan pola perilaku pelanggan.

Pada penelitian ini, pelanggan akan disegmentasi dengan algoritma *k-means* menurut perilaku pembayaran pelanggan agar dapat diukur tingkat potensial pelanggan. Setiap segmen pelanggan akan di klasifikasikan menurut kelayakan pelanggan dengan algoritma C4.5. Dengan begitu akurasi algoritma C4.5 akan lebih baik serta lebih cocok untuk diterapkan menurut tingkat potensial pelanggan.

METODOLOGI PENELITIAN

a. Analisa Kebutuhan

Data yang digunakan dalam penelitian ini merupakan data primer yang diambil dari data perusahaan penyedia layanan jasa. Observasi dilakukan pada perusahaan tersebut untuk mengumpulkan data pembayaran pelanggan aktif yang berlangganan TV kabel, maupun internet.

b. Perancangan Penelitian

Penelitian ini bertujuan untuk memprediksi kelayakan pelanggan pada perusahaan layanan jasa di Jakarta yang menyediakan layanan jasa dalam bentuk internet dan TV kabel juga mengalami masalah dengan adanya pelanggan yang menolak melakukan pembayaran atas jasa yang telah digunakan, oleh karena itu perlu tindakan dan kebijakan yang dapat diambil untuk menanggulangi dan mengurangi jumlah pelanggan yang tidak layak. Metode penelitian pada penelitian ini adalah penelitian eksperimen dengan tahapan penelitian seperti berikut:

1. Pengumpulan data

Pengumpulan data lebih dari sekedar mengambil data yang ada tetapi harus mampu mendeskripsikan data yang ada, serta memiliki kontribusi terhadap pengetahuan. Data tersebut harus jelas, memiliki relasi, dapat diukur, dapat diprediksi, memiliki generalisasi serta teori. Berdasarkan jenisnya data dibedakan menjadi:

- Data Sekunder yang merupakan data yang digunakan dari penelitian orang lain. Walaupun

dalam penelitian orang lain dibuat dengan tujuan berbeda namun dapat dimanfaatkan, data ini sangat mudah diperoleh dan cepat karena tersedia di tempat tempat umum seperti perpustakaan, toko buku, maupun internet

- Data Primer adalah data yang dikumpulkan langsung dari sumber data. Pengumpulan data ini memerlukan waktu serta biaya yang lebih banyak dari data sekunder. Contoh data primer adalah kuisioner, wawancara, hasil observasi dan eksperimen.

2. Pengolahan awal data

Data yang sudah dikumpulkan diolah dengan algoritma soft-computing untuk mengurangi data yang tidak relevan, atau data dengan atribut yang hilang. Pengolahan juga dapat berupa konversi nilai nilai redundan atau nilai yang terlalu beragam kedalam kelompok yang lebih kecil untuk mempermudah pembentukan model.

3. Model/metode yang diusulkan

Menggambarkan alur metode yang diusulkan serta menjelaskan cara kerja model yang diusulkan. Metode ini akan digambarkan secara skematik dan disertai dengan formula perhitungan. Model akan dibentuk dari data yang sudah diolah, dan hasil pengolahan model akan diukur dengan model yang ada saat ini.

4. Eksperimen dan pengujian model

Menjabarkan bagaimana eksperimen yang dilakukan hingga terbentuknya model, serta menjelaskan cara menguji model yang terbentuk.

5. Evaluasi dan validasi hasil

Evaluasi dilakukan dengan mengamati hasil prediksi menggunakan Algoritma Soft computing. Validasi dilakukan dengan mengukur hasil prediksi dibandingkan dengan data asal. Pengukuran kinerja dilakukan dengan membandingkan nilai error hasil prediksi masing-masing algoritma sehingga dapat diketahui algoritma yang lebih akurat.

HASIL DAN PEMBAHASAN

a. Segmentasi Potensial Pelanggan

Data yang sudah ada akan disegmentasi dengan algoritma k-means, dengan atribut yang akan digunakan berupa tagihan 6 bulan kebelakang, saldo pelanggan sampai 6 bulan kebelakang, lama hari terakhir membayar dan umur pelanggan.

Seluruh nilai data distandarisasi dengan skala *min-max*. Dari keseluruhan data yang ada, diambil nilai minimal dan maksimal, setelah itu setiap data

diskalakan dengan minimal dan maksimal tersebut. Karena perbandingan keseluruhan rate dan balance dengan lastpay dan dan custage harus 1:1:1 sesuai dengan teori RFM, maka hasil skala untuk rate dan balance dikalikan seratus, sedangkan skala lastpay dan cust age dikalikan dengan 1400.

Tabel 1 Data yang akan disegmentasi dengan *k-means*

No	Rate Balance	LastPay	CustAge
1	223	1356	971
2	269	1362	1400
3	244	1367	1222
4	254	394	292
5	264	835	459
6	206	1362	1400
7	230	1368	867
8	185	1399	365
9	195	1399	605
10	681	1118	167

Sumber: Data Olahan(2018)

Secara acak akan dipilih data sejumlah dengan nilai k yaitu 5. Maka dipilih data nomor 1, 3, 5, 6, dan 8 sebagai pusat segmentasi pertama. Nilai *euclidean* diperhitungkan untuk setiap data kedalam ketiga titik tersebut.

Tabel 2 Hasil perhitungan titik euclidean pertama

Data	<i>centroi</i>				
No.	1	3	5	6	8
1	0.00	252.09	731.63	429.38	608.73
2	431.49	179.85	1078.5	62.98	1039.0
3	252.09	0.00	930.38	182.03	859.63
4	1177.8	1346.0	471.67	1472.0	1010.0
5	731.63	930.38	0.00	1080.0	577.28
6	429.38	182.03	1080.0	0.00	1035.8
7	104.91	355.27	672.11	533.57	504.98
8	608.73	859.63	577.28	1035.8	0.00
9	369.59	619.76	586.70	795.94	240.22
10	955.44	1168.8	582.38	1343.7	603.82

Sumber: Data Olahan (2018)

Tabel 3. Tabel Statistik Keseluruhan Data Untuk Segmentasi

Name	Type	Statistics	Range
rateNOW	numeric	avg = 7.694 +/- 5.387	[0.000 ; 61.528]
Rate6	numeric	avg = 6.964 +/- 4.565	[0.000 ; 52.180]
Rate5	numeric	avg = 7.051 +/- 4.721	[0.000 ; 60.106]
Rate4	numeric	avg = 7.111 +/- 4.821	[0.000 ; 60.238]
Rate3	numeric	avg = 7.443 +/- 5.187	[0.000 ; 61.133]
Rate2	numeric	avg = 7.464 +/- 5.171	[0.000 ; 61.133]
Rate1	numeric	avg = 7.546 +/- 5.223	[0.000 ; 61.133]
LastPay	numeric	avg = 1366.464 +/- 71.615	[394.000 ; 1400.000]
CustAge	numeric	avg = 164.786 +/- 244.323	[10.000 ; 1400.000]
BalanceNOW	numeric	avg = 5.160 +/- 6.291	[0.000 ; 97.073]
Balance6	numeric	avg = 34.908 +/- 2.744	[12.432 ; 74.563]
Balance5	numeric	avg = 31.230 +/- 3.141	[4.300 ; 81.197]
Balance4	numeric	avg = 27.283 +/- 3.563	[4.810 ; 76.089]
Balance3	numeric	avg = 25.488 +/- 3.892	[9.065 ; 83.539]
Balance2	numeric	avg = 18.689 +/- 4.451	[7.941 ; 80.391]
Balance1	numeric	avg = 40.663 +/- 3.528	[34.063 ; 92.016]

Sumber: Data Olahan(2018)

Tabel 4. Nilai Titik Tengah Setiap Segmen

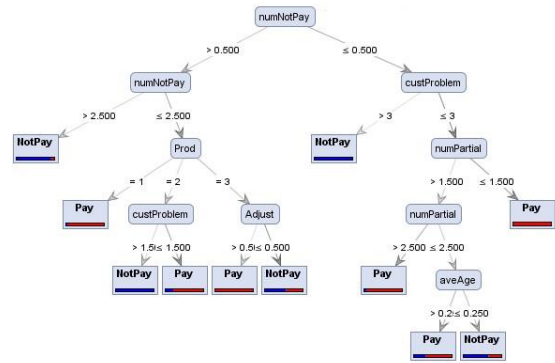
Attribute	cluster_0	cluster_1	cluster_2	cluster_3	cluster_4
rateNOW	7.594367	4.356739	5.67174	7.852482	8.126008
BalanceNOW	4.585035	24.24704	3.818727	4.293624	4.566758
Rate1	7.321036	4.320701	5.635567	7.772679	7.921956
Rate2	7.23015	4.323026	5.593635	7.678887	7.834718
Rate3	7.078448	4.338526	5.650754	7.74295	7.757297
Rate4	6.746599	4.165825	5.433084	7.298033	7.467799
Rate5	6.714208	4.503069	5.493992	7.180029	7.398141
Rate6	6.619499	4.994949	5.364581	7.084119	7.280499
Balance1	40.08429	51.43417	39.90542	40.17373	40.34714
Balance2	18.02442	31.78835	17.81404	18.1056	18.28969
Balance3	25.05501	35.96472	24.81821	25.1898	25.05136
Balance4	26.62417	35.98502	26.72197	27.02881	26.94768
Balance5	30.9311	38.18569	30.55372	31.0827	30.94078
Balance6	34.57337	40.18767	34.48082	34.76364	34.70558
LastPay	1378.678	1056.265	1370.94	1379.748	1378.826
CustAge	1167.819	171.7647	516.8503	149.2604	49.47193
Total Data	177	170	334	1490	2369

Sumber: Data Olahan(2018)

b. Klasifikasi Kelayakan Pelanggan

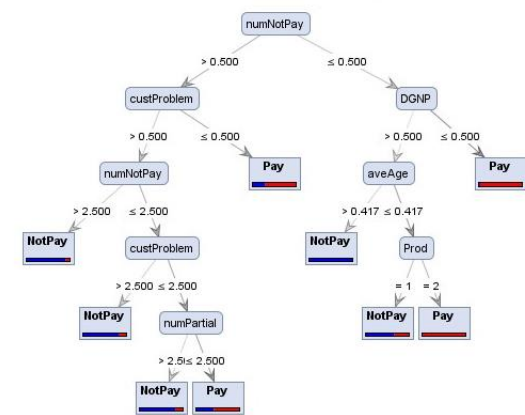
Setelah proses segmentasi, didapatkan 5 hasil segmentasi, segmen yang terbentuk dijadikan sebagai atribut baru untuk memudahkan pengolahan data dengan algoritma C4.5. Atribut atribut yang digunakan untuk proses segmentasi tidak lagi digunakan untuk klasifikasi kelayakan. Dengan begitu atribut yang tersisa yang akan digunakan untuk klasifikasi kelayakan pelanggan adalah adjustment, jumlah pelanggan tidak membayar dalam 6 bulan kebelakang, jumlah pelanggan membayar partial dalam 6 bulan kebelakang.

Setelah didapat segmen pelanggan, data setiap segmen digunakan untuk diklasifikasikan kelayakan pelanggan. Atribut yang dipilih merupakan atribut yang sudah dikurangi dengan atribut yang digunakan dalam segmentasi, adjustment, jumlah tidak membayar dan jumlah membayar partial dalam 6 bulan terakhir, rata-rata, maksimal dan minimal hutang dalam 6 bulan terakhir, jenis produk yang digunakan, jumlah dihubungi, serta pernahkan pelanggan downgrade atau disconnect.



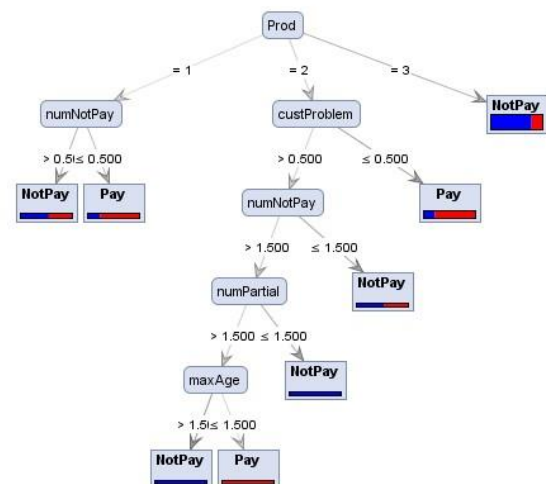
Sumber: Data Olahan(2018)

Gambar 1. Pohon Keputusan yang terbentuk untuk segmen pertama



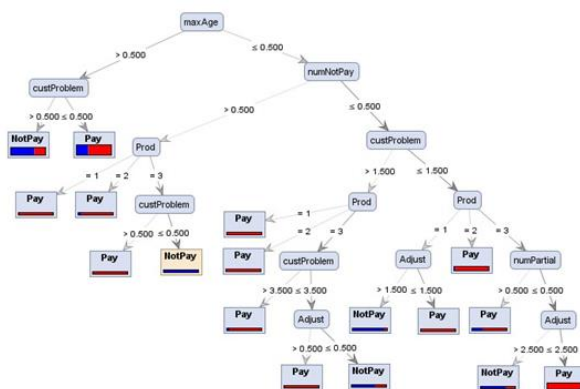
Sumber: Data Olahan(2018)

Gambar 2. Pohon Keputusan yang terbentuk untuk segmen kedua



Sumber: Data Olahan(2018)

Gambar 3. Pohon Keputusan yang terbentuk untuk segmen ketiga



Sumber: Data Olahan(2018)

Gambar 4. Pohon Keputusan yang terbentuk untuk segmen keempat

Jika seorang pelanggan A memiliki maksimal hutang 1 dan tidak pernah complain (artinya custProblem berisi 0) dan tergolong dalam segmen 4. Maka pelanggan tersebut akan diberikan kelas "PAY" yang berarti pelanggan tersebut merupakan pelanggan layak. Hal ini sesuai dengan aturan pertama pada pohon yang terbentuk untuk segmen pelanggan keempat. Semua data pelanggan yang berumur lebih dari 6 bulan dapat diprediksi kelayakannya, karena model yang terbentuk membutuhkan data lebih dari 6 bulan untuk dapat bekerja.

Sebagai ukuran pengujian, data yang sama tanpa atribut segmentasi pelanggan juga diterapkan dengan algoritma C4.5 dengan parameter yang sama dengan data yang sudah disegmentasi. Proses pengujian juga disamakan yaitu dengan *10 folds cross validation*. Dengan mendapatkan model yang baik, perusahaan dapat menerapkan model yang terbentuk langsung pada data pelanggan yang ada.

c. Evaluasi model *Confusion Matrix*

Confusion Matrix menampilkan hasil prediksi dalam bentuk tabel secara lengkap, hasil prediksi didapat dari rata-rata penerapan model yang terbentuk kedalam data testing yang dipilih dengan menggunakan *cross validation*. Model yang akan dibandingkan adalah model yang dibentuk dengan menggunakan algoritma C4.5 pada dataset yang sudah disegmentasi yang belum disegmentasi.

Pada dataset yang tidak disegmentasi, dataset terdiri dari atribut rateNow, rate1, rate2, rate3, rate4, rate5, rate6, balanceNow, bal1, bal2, bal3, bal4, bal5, bal6, Adjust, num not pay, num partial pay, max age, min age, ave age, DGNP, DINP, produk, cust problem, umur pelanggan, pembayaran terakhir. Kelas yang diprediksi merupakan kelas Pay yaitu pelanggan yang melakukan pembayaran tepat waktu, dan kelas notpay yaitu pelanggan yang menolak membayar, atau tidak melakukan pembayaran tepat waktu.

Dari hasil pengujian indikator, terlihat bahwa penerapan algoritma k-means dalam segmentasi pelanggan meningkatkan akurasi model yang cukup besar bila dibandingkan dengan data sebelum segmentasi. Hasil akurasi dari seluruh nilai *gain* dan *prepruning* yang diujikan menunjukkan peningkatan yang cukup besar. Namun indikator terbaik yang dipilih adalah dengan nilai *minimal gain* 0.06 dan *prepruning*. Walaupun akurasi yang terbentuk lebih rendah, namun model lebih sederhana dan memiliki AUC yang sebanding.

Confusion matrix dapat dilihat pada Tabel 5. Dengan jumlah prediksi notpay yang tepat berjumlah 235, sedangkan jumlah prediksi notpay yang ternyata membayar adalah 114 pelanggan. Sedangkan pelanggan yang diprediksi layak atau memiliki kelas pay, 1746 ternyata tidak melakukan pembayaran dan hanya 2444 pelanggan yang diprediksi layak dan benar benar melakukan pembayaran.

Akurasi model dihitung dari prediksi positif yang benar ditambah prediksi negatif yang benar dibandingkan dengan keseluruhan data. Akurasi model yang terbentuk untuk model yang tidak disegmentasi cukup rendah yaitu sebesar 59.02%.

Tabel 5. Confusion matrix untuk dataset yang tidak disegmentasi

accuracy: 59.02% +/- 3.37% (mikro: 59.02%)			
	true NotPay	true Pay	class precision
pred. NotPay	235	114	67.34%
pred. Pay	1746	2444	58.33%
class recall	11.86%	95.54%	

Sumber: Data Olahan(2018)

Dengan menggunakan algoritma k-means, atribut saldo, tagihan, umur pelanggan, dan tanggal pembayaran terakhir dikelompokkan menjadi atribut potensial. Dengan adanya atribut potensial, atribut lain yang masih digunakan adalah adjust, num notpay, num partial pay, max age, min age, ave age, DGNP, DISNP, produk, dan customer problem.

Confusion matrix dapat dilihat pada Tabel 5, dimana pelanggan yang diprediksi tidak layak dan memang ternyata tidak layak berjumlah 1464 pelanggan. Untuk pelanggan tidak layak yang ternyata layak berjumlah 513 pelanggan. Sedangkan untuk pelanggan yang diprediksi layak dan ternyata tidak layak berjumlah 517 pelanggan. Dan pelanggan yang diprediksi layak dan ternyata memang layak berjumlah 2045 pelanggan.

Tabel 6. Confusion Matrix untuk data setelah segmentasi

accuracy: 77.31% +/- 2.09% (mikro: 77.31%)

	true NotPay	true Pay	class precision
pred. NotPay	1464	513	74.05%
pred. Pay	517	2045	79.82%
class recall	73.90%	79.95%	

Sumber: Data Olahan(2018)

Akurasi keseluruhan model yang terbentuk setelah segmentasi adalah 77.31%. Dibandingkan dengan dataset yang belum disegmentasi, terdapat peningkatan akurasi yang cukup tinggi yaitu sebesar 18.29%.

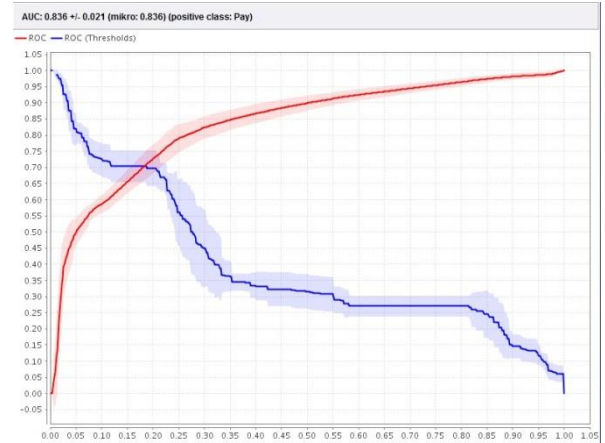
d. Kurva ROC (Receiver Operating Characteristic). Selain menggunakan Confusion matrix, evaluasi juga dilakukan dengan menggunakan *ROC Curve*. Pada dataset yang belum disegmentasi, nilai AUC yang didapat dari ROC Curve adalah sebesar 0.537.



Sumber: Data Olahan(2018)

Gambar 5. ROC Curve untuk dataset sebelum segmentasi

Sedangkan pada dataset yang sudah disegmentasi nilai AUC mengalami peningkatan hingga nilai AUC menjadi 0.836. Dari peningkatan akurasi dan nilai AUC model, terlihat bahwa pengolahan awal dataset dengan menggunakan k-means dapat menghasilkan model yang lebih baik. Peningkatan yang didapat cukup tinggi karena nilai akurasi meningkat hingga 18.29% dan nilai AUC meningkat menjadi 0.836.



Sumber: Data Olahan(2018)

Gambar 6. ROC Curve untuk dataset setelah segmentasi

d.Implikasi Penelitian

Klasifikasi data dengan algoritma C4.5 sangat baik jika didukung dengan dataset yang baik. Untuk klasifikasi kelayakan, salah satu attribut yang berpengaruh adalah keuangan pelanggan. Sedangkan pada perusahaan layanan jasa, data keuangan pelanggan yang dapat diperoleh sangat minim. Oleh karena itu semua attribut yang dapat dijadikan acuan untuk menentukan potensial pelanggan perlu dikumpulkan.

Tabel 7. Perbandingan model C4.5 sebelum dan sesudah segmentasi

	C4.5	C4.5 + k-means
Attribut	26 attribut	11 attribut (16 untuk k-means)
Akurasi	59.02%	77.31%
AUC	0.537	0.836

Sumber: Data Olahan(2018)

Penggunaan attribut yang terlalu banyak akan mengurangi proses klasifikasi dan menurunkan akurasi. Attribut dengan nilai informasi gain yang terlalu kecil akan membuat hasil pohon keputusan yang terbentuk menjadi kompleks, dan tingkat akurasi yang rendah. Terlalu banyak atribut numerik juga akan membuat pohon yang terbentuk memiliki banyak cabang yang terduplikasi.

Tabel 8. Segmen pelanggan yang terbentuk

Attribute	Segmen 1	Segmen 2	Segmen 3	Segmen 4	Segmen 5
Balance + Rate	229.1817	288.7955	216.9563	233.2471	234.6354
LastPay	1378.678	1056.265	1370.94	1379.748	1378.826
CustAge	1167.819	171.7647	516.8503	149.2604	49.47193
Total Data	177	170	334	1490	2369

Sumber: Data Olahan(2018)

Segmentasi pelanggan dibentuk dengan faktor RFM, menggambarkan persebaran pelanggan sesuai dengan faktor yang dipilih, dimana jumlah dari *balance* dan *rate* digabungkan menjadi faktor

monetary, Lastpay yang terdiri dari tanggal pelanggan melakukan pembayaran dianggap sebagai faktor *recency*. Dan umur pelanggan dipilih menjadi faktor *frequency* karena semakin tinggi umur pelanggan, maka semakin sering pula pelanggan melakukan pembayaran.

Segmen 1 (177 pelanggan) memiliki rata-rata yang tinggi dalam ketiga faktor, oleh karena itu masuk pada potensial sangat tinggi. Segmen 2 (170 pelanggan) hanya memiliki *recency* yang tinggi, dikategorikan sebagai pelanggan dengan tingkat potensial sangat rendah. Segmen 3 (334 pelanggan) memiliki *monetary* paling rendah dan dikategorikan sebagai potensial rendah. Segmen 4 (1490 pelanggan) memiliki tagihan dan *recency* yang tinggi, termasuk potensial tinggi. Dan segmen terakhir (2369 pelanggan) tergolong potensial menengah dengan *monetary* tinggi, namun *frequency* yang rendah.

Proses segmentasi untuk mengelompokkan beberapa atribut yang numerik dapat membantu membuat atribut baru serta mengurangi atribut sehingga mampu meningkatkan akurasi C4.5. Dengan proses segmentasi, terlihat bahwa akurasi dari proses klasifikasi meningkat dari 59.02 % menjadi 77.31%. Selain itu AUC juga meningkat dari 0.537 menjadi 0.836. Selain itu segmen pelanggan juga merupakan salah satu kebutuhan perusahaan agar dapat mengenal pelanggannya, sehingga pendekatan terhadap pelanggan tidak layak, bahkan promosi perusahaan dapat disesuaikan dengan segmen pelanggan.

KESIMPULAN

Dari penelitian yang dilakukan, pengurangan dimensi atribut dalam proses klasifikasi kelayakan pelanggan terbukti dapat meningkatkan akurasi model yang terbentuk. Pada perusahaan layanan multimedia, atribut dapat dikelompokkan dengan algoritma data mining seperti *k-means*. Pengelompokan attribute atau yang disebut dengan ekstraksi fitur ini sangat efektif untuk mengurangi dimensi data sekaligus membentuk atribut yang dapat diandalkan.

Model yang terbentuk dengan algoritma C4.5 pada atribut data yang dikelompokkan dengan *k-means* menunjukkan peningkatan yang signifikan bila dibandingkan dengan data yang atributnya tidak dikelompokkan. Perusahaan dapat lebih baik dalam membedakan pelanggan yang layak dan tidak layak. Langkah-langkah yang diambil untuk menghadapi pelanggan yang tidak layak pun dapat disesuaikan dengan tingkat potensial dari pelanggan itu sendiri.

Peningkatan kualitas model dapat dilihat dari peningkatan akurasi yang diukur dengan menggunakan confusion matrix, akurasi untuk model algoritma C4.5 yang tidak disegmentasi adalah sebesar 59.02 % dan AUC sebesar 0.537.

Setelah dilakukan ekstraksi fitur dengan *k-means* nilai akurasi meningkat menjadi

77.31% dan nilai AUC menjadi 0.836. Dengan menggunakan fitur ekstraksi, dimensi data dapat dikurangi, sekaligus atribut yang terbentuk dapat lebih mewakili data yang diuji sehingga mampu menyederhanakan model yang terbentuk, sekaligus meningkatkan akurasi model.

Model yang terbentuk dapat diterapkan pada seluruh data pelanggan (dengan atribut yang mencukupi) sehingga perusahaan dapat langsung melihat pelanggan yang layak dan tidak layak. Pengenalan tingkat kelayakan pelanggan membantu perusahaan menyusun strategi pemasaran, serta dapat mengurangi beban untuk mempertahankan layanan terhadap pelanggan yang tidak layak.

REFERENSI

- Ali, S. A., Sulaiman, N., Mustapha, A., & Mustapha, N. (2009). K-means clustering to improve the accuracy of Decision tree response classification. *Information Technology Journal*.
<https://doi.org/10.3923/itj.2009.1256.1262>
- Daskalaki, S., Kopanas, I., Goudara, M., & Avouris, N. (2003). Data mining for decision support on customer insolvency in telecommunications business - Daskalaki et al. - 2003.pdf, 145, 239–255.
- Liao, T. W., & Triantaphyllou, E. (n.d.). *Recent advances in data mining of enterprise data: algorithms and applications*.
- Liu, Y., & Schumann, M. (n.d.). Data mining feature selection for credit scoring models.
- Pinheiro, C. A. R., Evsukoff, A. G., & Ebecken, N. F. F. (2006). Revenue recovering with insolvency prevention on a Brazilian telecom operator. *ACM SIGKDD Explorations Newsletter*, 8(1), 65–70.
<https://doi.org/10.1145/1147234.1147244>
- Prasad, P. (2011). Generating Customer Profiles for Retail Stores Using Clustering Techniques. *International Journal on Computer Science and Engineering*, 3(6), 2506–2510. Retrieved from
http://www.doaj.org/doaj?func=fulltext&passMe=http://www.enggjournals.com/ijcse/doc/IJ_CSE11-03-06-141.pdf
- Putra, R. R., & Wadisman, C. (2018). IMPLEMENTASI DATA MINING PEMILIHAN PELANGGAN POTENSIAL MENGGUNAKAN ALGORITMA K-MEANS, 1(1), 72–77.
- Sari, R. D. I., & Sindunata, Y. (2014). Penerapan Data Mining Untuk Analisa Pola Perilaku, 8(2), 10–16.
- Thiruvadi, S., & Patel, S. C. (n.d.). Survey of data-mining techniques used in fraud detection and

- prevention.
- Ali, S. A., Sulaiman, N., Mustapha, A., & Mustapha, N. (2009). K-means clustering to improve the accuracy of Decision tree response classification. *Information Technology Journal*. <https://doi.org/10.3923/itj.2009.1256.1262>
- Daskalaki, S., Kopanas, I., Goudara, M., & Avouris, N. (2003). Data mining for decision support on customer insolvency in telecommunications business - Daskalaki et al. - 2003.pdf, 145, 239–255.
- Liao, T. W., & Triantaphyllou, E. (n.d.). *Recent advances in data mining of enterprise data: algorithms and applications*.
- Liu, Y., & Schumann, M. (n.d.). Data mining feature selection for credit scoring models.
- Pinheiro, C. A. R., Evsukoff, A. G., & Ebecken, N. F. F. (2006). Revenue recovering with insolvency prevention on a Brazilian telecom operator. *ACM SIGKDD Explorations Newsletter*, 8(1), 65–70. <https://doi.org/10.1145/1147234.1147244>
- Prasad, P. (2011). Generating Customer Profiles for Retail Stores Using Clustering Techniques. *International Journal on Computer Science and Engineering*, 3(6), 2506–2510. Retrieved from http://www.doaj.org/doi/func=fulltext&passMe=http://www.enggjournals.com/ijcse/doc/IJ_CSE11-03-06-141.pdf
- Putra, R. R., & Wadisman, C. (2018). IMPLEMENTASI DATA MINING PEMILIHAN PELANGGAN POTENSIAL MENGGUNAKAN ALGORITMA K-MEANS, 1(1), 72–77.
- Sari, R. D. I., & Sindunata, Y. (2014). Penerapan Data Mining Untuk Analisa Pola Perilaku, 8(2), 10–16.
- Thiruvadi, S., & Patel, S. C. (n.d.). Survey of data-mining techniques used in fraud detection and prevention.

PROFIL PENULIS

Ela Nurelasari, M.Kom. Lahir di Ciamis 24 Desember 1987, Tahun 2010 lulus Program Diploma Tiga (DIII) dari AMIK BSI Jakarta Program Studi Komputerisasi Akuntansi. Tahun 2012 Lulus Strata Satu (SI) dari STMIK Nusa Mandiri Jakarta Jurusan Sistem Informasi dan Tahun 2015 lulus Strata Dua (S2) dari Pasca Sarjana Magister Ilmu Komputer STMIK Nusa Mandiri Jakarta. Menjadi Dosen tetap dari tahun 2013 sampai sekarang pada program studi Manajemen Informatika AMIK BSI Jakarta. Jabatan Fungsional Dosen sebagai Asisten Ahli dari awal tahun 2017 dan mendapatkan Sertifikasi Dosen pada tahun 2018. Menulis paper ilmiah diberbagai jurnal ilmiah baik di dalam kampus ataupun di luar kampus.