

METODE ANALISIS DISKRIMINAN KUADRAT TERKECIL PARSIAL UNTUK KLASIFIKASI SEGMENT LOYALITAS KONSUMEN SUSU PERTUMBUHAN*

Herdina Kuswari¹, Farit Mochamad Afendi^{2‡}, and Khairil Anwar Notodiputro³

¹Department of Statistics, IPB University, Indonesia, herdinakuswari@gmail.com

²Department of Statistics, IPB University, Indonesia, fmafendi@gmail.com

³Department of Statistics, IPB University, Indonesia, khairil@apps.ipb.ac.id

‡corresponding author

Indonesian Journal of Statistics and Its Applications (eISSN:2599-0802)

Vol 4 No 2 (2020), 284 - 295

Copyright © 2020 Herdina Kuswari, Farit Mochamad Afendi, and Khairil Anwar Notodiputro. This is an open-access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Consumer segmentation is the process of dividing consumers into different segments based on consumer characteristics, making it easier for companies to develop marketing strategies. The segmentation is carried out based on consumer loyalty using the RFM (Recency, Frequency, Monetary) approach a number of 7753 members of a nutritional product loyalty program is considered in the analysis. Partial least square discriminant analysis classification modeling is built using the results of consumer segmentation being the a response variable. The model is not good enough based on the AUC (Area Under Curve) value of the ROC (Relative Operating Characteristic) curve that quite low for each segment. The explanatory variables that have high contribution to the model is X5, X9, and X2 with VIP (Variable Importance in the Projection) values more than 1.

Keywords: classification, consumer segmentation, partial least square discriminant analysis, RFM.

*Received Nov 2019; Accepted Jun 2020; Published online on Jul 2020

1 Pendahuluan

Sektor bisnis sudah berkembang sangat pesat akibat pengaruh dari perkembangan teknologi informasi. Alasan tersebut mengakibatkan munculnya perusahaan pesaing yang menjadi tantangan baru bagi perusahaan lainnya. Konsumen merupakan salah satu bagian penting dalam perusahaan. Hal ini dikarenakan prospek bisnis yang baik pasti memiliki konsumen yang puas dan loyal terhadap perusahaan. Strategi meningkatkan dan mempertahankan konsumen harus dilakukan agar konsumen tidak beralih memilih produk dari perusahaan pesaingnya.

Customer Relationship Management (CRM) merupakan strategi untuk mengelola interaksi antara konsumen dengan perusahaan. Fase pertama dalam CRM adalah identifikasi konsumen, fase ini melibatkan penargetan populasi yang paling mungkin menjadi konsumen atau yang paling menguntungkan bagi perusahaan (Ngai et al., 2009). Salah satu elemen dalam identifikasi konsumen adalah segmentasi konsumen. Menurut Tsiptsis dan Chorianopoulos (2009), segmentasi konsumen memungkinkan perusahaan untuk memahami karakteristik konsumen pada masing-masing segmen sehingga memudahkan perusahaan membangun strategi pemasaran yang sesuai dengan karakteristiknya. Salah satu pendekatan yang umum digunakan untuk membentuk segmen konsumen adalah *Recency, Frequency* atau frekuensi, *Monetary* atau keuangan (RFM). RFM merupakan sebuah metode dalam mengkategorisasi konsumen berdasarkan peubah *recency*, frekuensi dan keuangan (Hughes, 1994). *Recency* menginformasikan waktu pembelian terakhir seorang konsumen. Konsumen yang paling baru melakukan pembelian berpeluang tinggi untuk melakukan pembelian ulang. Frekuensi adalah jumlah transaksi yang dilakukan konsumen dalam jangka waktu tertentu sedangkan keuangan adalah jumlah uang yang dikeluarkan oleh konsumen dalam jangka waktu tertentu untuk membeli produk tersebut. Penelitian sebelumnya mengenai segmentasi konsumen dengan pendekatan RFM dilakukan oleh Yani (2013) untuk menganalisa segmen konsumen yang potensial pada maskapai penerbangan Garuda dengan menggunakan metode k-rataan dan menghasilkan tujuh segmen konsumen

Perusahaan melakukan evaluasi konsumen sekali setiap periode yang telah ditentukan. Pembentukan segmen konsumen mempermudah perusahaan untuk membentuk metode pemasaran berdasarkan karakteristik konsumen per segmen yang terbentuk. Evaluasi yang dilakukan setiap periode akan memberikan gambaran perkembangan konsumen lama. Selain itu, konsumen baru juga akan masuk ke segmen yang telah terbentuk sebelumnya. Pembentukan segmen dari konsumen baru dilakukan berdasarkan model yang sudah dibuat dari model klasifikasi konsumen lama dengan peubah penjas yang mudah diamati.

Metode Analisis Diskriminan Kuadrat Terkecil Parsial merupakan salah satu metode klasifikasi yang banyak dilakukan dalam bidang kemometrika yang didasarkan pada pendekatan Kuadrat Terkecil Parsial (KTP) (Barker and Rayens, 2003). Analisis Diskriminan Kuadrat Terkecil Parsial terdiri dari regresi KTP klasik yang peubah responnya bersifat kategorik yang menunjukkan suatu kelas klasifikasi (Galtier et al., 2011). Metode ini dapat digunakan untuk menganalisis data yang memiliki korelasi yang tinggi, memiliki peubah penjas yang sangat banyak, dan juga secara simultan memodelkan beberapa peubah respon (Wold et al., 2001).

Pada penelitian ini dilakukan segmentasi konsumen susu pertumbuhan dengan model RFM. Segmen akan dibentuk berdasarkan loyalitas konsumen menggunakan

Tabel 1: Peubah penyusun segmentasi konsumen

Peubah	Keterangan	Skala
R	Rentang waktu transaksi terakhir konsumen dengan tanggal penghimpunan data	Numerik
F	Rata-rata transaksi konsumen perbulan selama menjadi anggota program loyalitas	Numerik
M	Rata-rata dalam gram produk yang dibeli konsumen perbulan selama menjadi anggota dibagi dengan umur anak dalam tahun	Numerik

pendekatan skor akhir dari penjumlahan skor peubah *recency*, *frequency* dan *monetary*. Identifikasi dari setiap segmen akan dilakukan menggunakan eksplorasi data terkait setiap segmen serta dengan klasifikasi. Metode klasifikasi yang akan digunakan yaitu Analisis Diskriminan Kuadrat Terkecil Parsial. Model klasifikasi yang terbentuk diharapkan dapat mempercepat penentuan segmen bagi konsumen baru serta mempertahankan konsumen lama.

Tujuan penelitian ini untuk melakukan segmentasi konsumen susu pertumbuhan menggunakan pendekatan peubah *Recency*, *Frequency*, dan *Monetary* (RFM) dengan metode klasifikasi analisis diskriminan kuadrat terkecil parsial.

2 Metodologi

2.1 Bahan dan Data

Data yang digunakan pada penelitian ini adalah data profil dan aktivitas anggota program loyalitas suatu produk nutrisi yang dihimpun pada tanggal 02 Juli 2018. Objek dari data tersebut ialah seluruh anggota di Indonesia yang terdaftar di tahun 2017 dan bertransaksi diantara tanggal 01 Januari sampai 01 Juli tahun 2018. Anggota program loyalitas diasumsikan adalah orang tua dari anak umur 12 sampai 60 bulan yang mengonsumsi susu pertumbuhan dan keanggotaannya dalam program tersebut bersifat sukarela. Data profil didapatkan ketika anggota pertama kali mendaftar pada program.

Segmentasi dilakukan pada satu merek yang terdiri dari 7753 amatan. Pendekatan RFM dilakukan berdasarkan peubah yang terdapat pada Tabel 1. Selanjutnya, dilakukan klasifikasi dengan hasil segmentasi konsumen sebagai peubah respon dan 24 peubah penjelas yang terdapat pada Tabel 2. Peubah penyusun untuk klasifikasi didapatkan dari data prol (X1-X7) dan data karakteristik provinsi dari Badan Pusat Statistik (BPS) (X8-X24). Setiap anggota yang memiliki asal provinsi yang sama akan memiliki nilai X2, X8 sampai dengan X24 sama. Data karakteristik provinsi ini diperoleh dari *website* BPS

2.2 Prosedur Analisis Data

Prosedur analisis data dilakukan menggunakan menggunakan perangkat lunak Microsoft Excel dan R 3.5.2 dengan menggunakan paket *Discriminer*, *ggplot2*, *corrplot*, *pROC*, dan *rgdal*. Adapun langkah-langkah dalam melakukan analisis yang digunakan dalam penelitian ini terbagi menjadi dua bagian.

Tabel 2: Peubah penyusun klasifikasi

Peubah	Keterangan	Skala
Y	Hasil Segmentasi (Respon)	Kategorik
X1	Lama keanggotaan konsumen (bulan)	Numerik
X2	Proporsi anggota program loyalitas setiap provinsi	Numerik
X3	Melakukan transaksi setelah registrasi di bulan yang sama	Kategorik
X4	Saluran yang digunakan konsumen untuk mengakses program loyalitas	Kategorik
X5	Umur anak (tahun)	Numerik
X6	Jumlah transaksi anggota program loyalitas pada bulan pertama	Numerik
X7	Jumlah produk yang dibeli oleh anggota bulan pertama (gram)	Numerik
X8	Laju pertumbuhan penduduk per provinsi	Numerik
X9	Proporsi pasangan usia subur per provinsi	Numerik
X10	Koefisien Gini provinsi	Numerik
X11	Indeks Pembangunan Manusia provinsi	Numerik
X12	Proporsi desa dengan minimarket per provinsi	Numerik
X13	Proporsi desa dengan toko kelontong per provinsi	Numerik
X14	Proporsi desa dengan pasar dengan bangunan per provinsi	Numerik
X15	Persentase penduduk miskin per provinsi	Numerik
X16	Proporsi rumah tangga berdasarkan status kepemilikan rumah sendiri menurut provinsi	Numerik
X17	Proporsi konsumsi protein sehari menurut provinsi	Numerik
X18	Proporsi bayi usia kurang dari 6 bulan yang mendapatkan asi eksklusif menurut provinsi	Numerik
X19	Proporsi remaja dan dewasa dengan keterampilan teknologi informasi dan komunikasi per provinsi	Numerik
X20	Proporsi individu yang menggunakan internet menurut Provinsi	Numerik
X21	Proporsi individu yang menggunakan telepon genggam	Numerik
X22	Proporsi balita kurus mendapat makanan tambahan per provinsi	Numerik
X23	Rerata kecukupan energi balita per provinsi	Numerik
X24	Proporsi balita gemuk per provinsi	Numerik

Segmentasi

Segmentasi dilakukan pada peubah dalam Tabel 1. Langkah-langkah segmentasi konsumen adalah sebagai berikut:

1. Melakukan eksplorasi data dengan statistika deskriptif pada masing-masing peubah untuk mengetahui gambaran konsumen susu pertumbuhan dalam program loyalitas tersebut.
2. Melakukan praproses data sebelum melakukan segmentasi konsumen dengan model RFM.
3. Membagi masing-masing peubah *recency*, *frequency* dan *monetary* (RFM) menjadi 5 bagian berdasarkan nilai kuantil sebaran datanya. Nilai setiap peubah ini dibagi rata 20% untuk setiap nilainya dengan mengurutkan data terlebih dahulu. Nilai *recency* 20% tertinggi akan mendapatkan nilai 1, 20% selanjutnya bernilai 2, dan seterusnya. Pembagian nilai *frequency* dan *monetary* berkebalikan dengan nilai *recency*. Nilai *frequency* dan *monetary* 20% tertinggi akan mendapat nilai 5, 20% selanjutnya akan mendapatkan nilai 4 dan seterusnya.
4. Menjumlahkan ketiga peubah tersebut untuk setiap konsumen menjadi skor akhir RFM.
5. Membagi skor akhir RFM menjadi 4 bagian yang sama besar berdasarkan nilai kuartil seluruh urutan data.

Klasifikasi

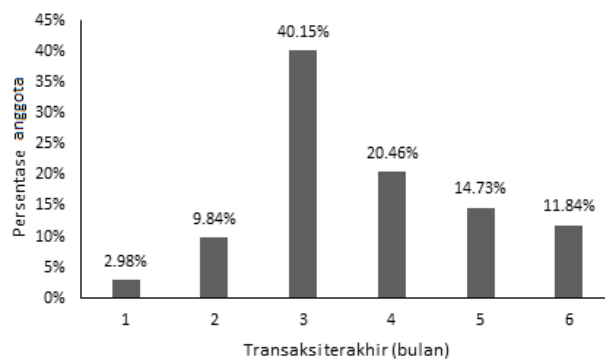
Klasifikasi dilakukan pada peubah dalam Tabel 2. Langkah-langkah pemodelan klasifikasi adalah sebagai berikut:

1. Eksplorasi hasil segmentasi yang diperoleh sehingga diperoleh karakteristik setiap segmen yang terbentuk.
2. Seleksi peubah penjelas dengan melihat korelasi antar peubah X dan nilai akar ciri dari matriks $X'X$.
3. Melakukan pemodelan klasifikasi Analisis Diskriminan Kuadrat Terkecil Parsial 10-*fold* validasi silang untuk keseluruhan data dengan jumlah komponen sebanyak peubah penjelas yang telah diseleksi. Validasi silang ini dilakukan untuk memilih jumlah komponen optimal yang akan digunakan. Komponen optimal dipilih berdasarkan nilai salah klasifikasi minimal dari seluruh komponen.
4. Melakukan proses Analisis Diskriminan Kuadrat Terkecil Parsial dengan jumlah komponen sebanyak komponen optimal yang telah dipilih.
5. Membuat plot dua dimensi antara skor X dua komponen pertama untuk melihat penyebaran data setelah proses Analisis Diskriminan Kuadrat Terkecil Parsial.
6. Menghitung nilai VIP untuk setiap peubah penjelas.
7. Menghitung kinerja klasifikasi menggunakan kurva ROC. Kinerja klasifikasi yang diukur dengan nilai AUC dan kemampuan memprediksi tiap segmen.

3 Hasil dan Pembahasan

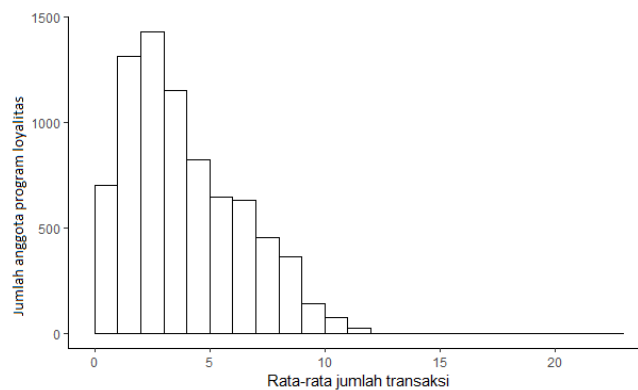
3.1 Eksplorasi Data

Data konsumen program loyalitas yang digunakan pada penelitian ini terdiri dari 7753 konsumen susu pertumbuhan merek X. Berdasarkan Gambar 1, waktu terakhir transaksi anggota paling banyak terjadi antara tiga bulan sebelum tanggal penghimpunan data yakni sejumlah 40.15%. Anggota program loyalitas paling sedikit melakukan transaksi terakhir sebulan sebelumnya yaitu hanya sejumlah 2.98%. Anggota dengan waktu transaksi terakhir satu bulan sebelumnya bisa disebut sebagai anggota yang lebih loyal, karena memiliki peluang untuk melakukan transaksi kembali. Rentang waktu transaksi terakhir dari anggota ini akan membentuk nilai *recency* sebagai salah satu komponen segmentasi loyalitas konsumen.



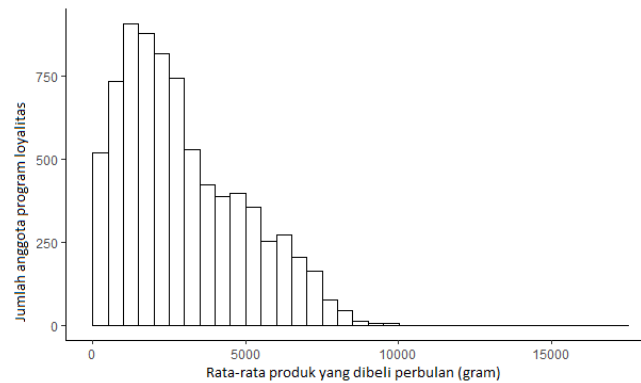
Gambar 1: Persentase waktu transaksi terakhir anggota program loyalitas dalam bulan

Rata-rata transaksi anggota dapat dilihat pada Gambar 2. Rata-rata transaksi anggota perbulan merupakan jumlah seluruh transaksi per anggota dibagi dengan lama keanggotaan dalam bulan. Rentang rata-rata transaksi anggota perbulan ini berada diantara 0 sampai 23 kali dalam sebulan, dengan sebaran rata-rata transaksi menjulur ke kanan. Rata-rata transaksi anggota terbanyak berada di antara satu sampai tiga kali transaksi dalam sebulan. Rata-rata frekuensi terbanyak dilakukan oleh dua anggota dengan jumlah transaksi antara 21 sampai 23 kali dalam sebulan. Hal ini dapat menyebabkan rata-rata dari rata-rata frekuensi transaksi anggota perbulan menjadi tinggi.



Gambar 2: Sebaran rata-rata transaksi anggota program loyalitas perbulan

Rata-rata produk yang dibeli oleh konsumen dalam gram dapat dilihat pada Gambar 3. Rata-rata produk yang dibeli oleh anggota perbulan merupakan total produk yang dibeli anggota dalam gram dibagi dengan lama keanggotaan dalam bulan. Rentang pembelian produk ini berada diantara 12.15 gram sampai dengan 17649.78 gram. Frekuensi anggota paling banyak membeli susu pertumbuhan adalah antara 500 gram sampai dengan 2500 gram dalam sebulan. Sebaran yang cenderung menjulur ke kanan menunjukkan terdapat beberapa anggota yang melakukan pembelian dalam jumlah gram yang tinggi dalam sebulan. Terdapat dua anggota yang melakukan pembelian susu pertumbuhan hingga lebih dari 12000 gram perbulannya.



Gambar 3: Sebaran rata-rata produk yang dibeli konsumen perbulan dalam gram

Konsumen dapat mengakses program loyalitas melalui tiga saluran. *Call Center*, SMS, dan WEB merupakan tiga saluran yang dapat digunakan. WEB merupakan saluran yang paling banyak digunakan oleh anggota dalam program loyalitas ini. Persentase anggota yang menggunakan saluran WEB sebanyak 88.48% dari total anggota. Sebanyak 11.51% anggota mengakses program loyalitas melalui *Call Center* dan SMS. Saluran yang digunakan oleh anggota dapat membantu dalam melakukan pendekatan dengan konsumen untuk pemasaran perusahaan.

Anggota dalam program loyalitas ini tersebar dalam 34 provinsi di Indonesia. Provinsi Jawa Timur merupakan provinsi dengan persentase anggota terbesar yakni sebesar 26.92%. Persentase terbesar selanjutnya adalah provinsi Jawa Barat, Jawa Tengah, DKI Jakarta dan Banten dengan persentase anggota berturut-turut 15.03%, 12.89%, 9.63%, 6.15% sebagai 5 provinsi dengan persentase anggota terbesar. Provinsi dengan jumlah anggota paling sedikit adalah provinsi Papua Barat dengan persentase anggota hanya sebesar 0.03%. Karakteristik provinsi di Indonesia yang cenderung berbeda akan digunakan dalam membentuk peubah penjelas dalam proses klasifikasi.

3.2 Segmentasi Konsumen dengan Model RFM

Segmentasi loyalitas konsumen dilakukan menggunakan skor akhir dari penjumlahan skor RFM. Skor R, F, dan M diperoleh dari pembagian nilai peubah R, F, M menjadi 5 skor berdasarkan Tabel 3. Skor ini berasal dari sebaran data asal masing-masing peubah yang dibagi 5 bagian sama besar setelah data diurutkan. Peubah *recency* merupakan peubah dengan rentang antara 1 sampai dengan 179. Semakin rendah nilai peubah *recency* dari anggota maka semakin tinggi skor R yang didapatkan. Hal ini berbanding terbalik dengan peubah frekuensi dan keuangan. Anggota dengan nilai frekuensi dan keuangan yang tinggi akan memiliki skor F dan M yang tinggi. Semakin tinggi skor dari RFM maka

Tabel 3: Pembagian skor untuk setiap peubah *Recency, Frequency, Monetary*

Skor R	R	Skor F	F	Skor M	M
1	130-179	1	0.05 - 1.66	1	3.83 - 450.31
2	101-129	2	1.66 - 2.73	2	450.31 - 834.90
3	79-100	3	2.73 - 4.06	3	834.90 - 1305.46
4	64-78	4	4.06 - 6.30	4	1305.46 - 1970.21
5	1-63	5	6.30 - 22.33	5	1970.21 - 7534.54

Tabel 4: Pembagian segmen konsumen berdasarkan sebaran nilai skor akhir

Skor Akhir	Segmen	Jumlah Anggota
13 - 15	1	1210
10 - 12	2	2496
7 - 9	3	2095
3 - 6	4	1952

diasumsikan bahwa anggota tersebut semakin loyal. Hal ini berpengaruh terhadap skor akhir yang didapatkan. Semakin tinggi skor akhir seorang anggota maka semakin tinggi tingkat loyalitasnya.

Skor akhir RFM dari seluruh anggota akan dibagi menjadi empat segmen berdasarkan nilai kuartil sebaran skor akhirnya (Tabel 4). Skor akhir RFM menunjukkan tingkat loyalitas konsumen, semakin tinggi nilainya maka akan semakin tinggi tingkat loyalitasnya. Segmen satu dengan jumlah nilai akhir RFM tertinggi yakni antara nilai 13 sampai 15 adalah segmen dengan tingkat loyalitas yang sangat loyal. Segmen dua merupakan segmen dengan selang nilai tertinggi kedua yakni 10 sampai 12. Segmen ini merupakan segmen dengan loyalitas loyal. Segmen tiga dan segmen empat berturut-turut memiliki selang nilai antara 7 sampai 9 dan 3 sampai 6 disebut dengan segmen tidak loyal dan segmen sangat tidak loyal.

3.3 Seleksi Peubah Penjelas Klasifikasi

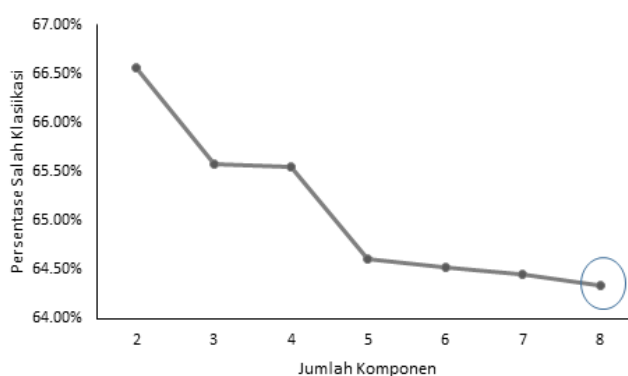
Beberapa peubah penjelas klasifikasi (Tabel 2) dalam penelitian ini berkorelasi tinggi antara peubah satu dengan peubah lainnya. Korelasi yang tinggi antar peubah penjelas tersebut menyebabkan informasi yang tumpang tindih sehingga tidak semua peubah penjelas harus digunakan dalam proses klasifikasi. Tumpang tindihnya informasi peubah penjelas tersebut dapat dilihat dari nilai akar ciri minimum matriks $X'X$ yang rendah.

Matriks $X'X$ merupakan cerminan dari matriks korelasi. Karena dalam penelitian ini juga digunakan peubah penjelas kategorik, maka agar $X'X$ dapat digunakan untuk menghitung akar ciri, semua peubah tersebut tidak diikutsertakan dalam perhitungan. Nilai minimum akar ciri dari matriks $X'X$ untuk semua peubah penjelas numerik dalam matriks X cukup rendah yakni 0.0158811. Nilai tersebut mengindikasikan bahwa keragaman matriks X cukup rendah dijelaskan oleh beberapa peubah penjelas. Susunan peubah penjelas yang dipilih merupakan peubah-peubah yang memiliki nilai akar ciri yang cukup tinggi dari matriks $X'X$ dan korelasi antar peubah yang cukup rendah. Peubah penjelas yang dipilih berjumlah 8 dengan 3 peubah merupakan profil anggota dan 5 peubah merupakan karakteristik provinsi asal anggota program loyalitas. Peubah tersebut adalah $X_1, X_2, X_5, X_8, X_9, X_{10}, X_{18}, X_{20}$ dengan akar ciri minimum matriks

X'X bernilai 5.239799.

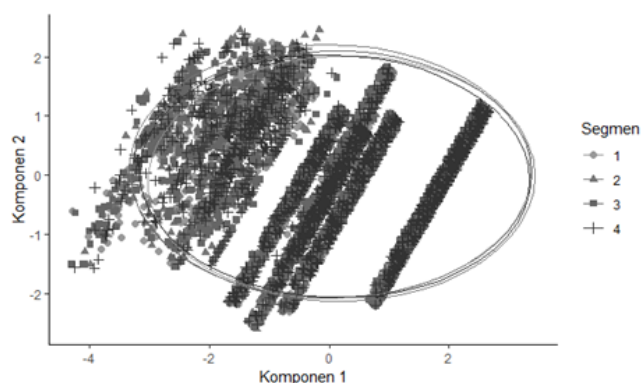
3.4 Klasifikasi dengan Metode Analisis Diskriminan Kuadrat Terkecil Parsial

Penelitian ini menggunakan 10-fold validasi silang untuk menentukan banyaknya komponen KTP. Komponen awal yang dibentuk sebanyak 8 komponen karena jumlah minimum dari peubah penjelas dan amatan adalah 8. Hasil yang didapatkan dari validasi silang ini adalah nilai salah klasifikasi untuk setiap komponen seperti Gambar 4. Hasil ini tidak cukup baik dikarenakan nilai salah klasifikasi yang besar yakni diatas 60%. Banyaknya komponen yang akan dipilih adalah nilai salah klasifikasi terkecil dari seluruh komponen. Nilai salah klasifikasi terkecil adalah 64.37% pada jumlah komponen 8. Proses klasifikasi dengan analisis diskriminan kuadrat terkecil parsial selanjutnya akan menggunakan jumlah komponen sebanyak 8 komponen.



Gambar 4: Nilai salah klasifikasi untuk setiap komponen KTP

Hasil skor X yang didapatkan dari analisis ini kemudian dibuat plot dua dimensi yaitupada dua skor komponen utama pertama yang diilustrasikan pada Gambar 5. Kontribusi dua skor X pertama tidak cukup baik dalam menjelaskan model yakni hanya sebesar 47% untuk kedua skor X. Skor X komponen pertama dan kedua memiliki kontribusi sebesar 33% dan skor X kedua kontribusi 14% terhadap model. Plot yang dihasilkanpun tidak cukup baik untuk memisahkan setiap segmen yang ada.

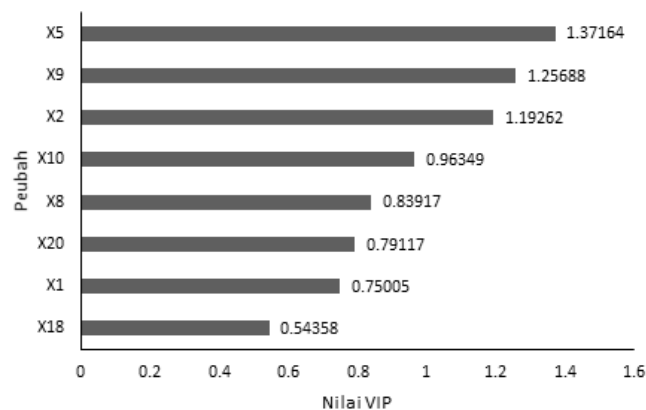


Gambar 5: Plot dua skor komponen utama pertama

Variable Importance in the Projection (VIP) merupakan salah satu teknik dan kriteria memilih peubah penjelas yang digunakan dalam analisis diskriminan KTP. Nilai VIP

juga dapat digunakan dalam melihat seberapa besar kontribusi peubah penjelas terhadap model. Semakin besar nilai VIP suatu peubah maka semakin besar kontribusi peubah tersebut dalam pemodelan klasifikasi.

Besarnya nilai VIP peubah penjelas dalam penelitian ini terdapat dalam Gambar 6. Penelitian ini menggunakan ambang batas peubah penjelas yang berkontribusi tinggi dalam model adalah peubah penjelas dengan nilai VIP lebih dari 1. Peubah dengan kontribusi tertinggi dalam model adalah X5 yakni umur anak dengan nilai VIP sebesar 1.37164. Peubah dengan nilai VIP lebih dari 1 selanjutnya adalah X9 yakni proporsi pasangan usia subur per provinsi dengan nilai VIP sebesar 1.25688 dan proporsi anggota setiap provinsi (X2) dengan nilai VIP sebesar 1.19262. Tiga peubah tersebut merupakan peubah yang memiliki kontribusi lebih besar dibanding peubah lainnya. Peubah dengan kontribusi paling kecil dalam pemodelan klasifikasi ini adalah proporsi bayi usia kurang dari 6 bulan yang mendapatkan asi eksklusif menurut provinsi (X18) dengan nilai VIP sebesar 0.54358



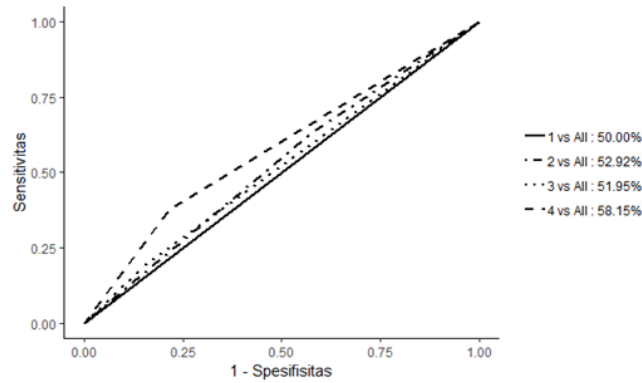
Gambar 6: Nilai VIP untuk setiap peubah penjelas dalam model

3.5 Evaluasi Model Analisis Diskriminan Kuadrat Terkecil Parsial

Kurva ROC digunakan sebagai dasar evaluasi model klasifikasi Analisis Diskriminan Kuadrat Terkecil Parsial. Nilai AUC (*Area Under Curve*) dari kurva ROC yang diperoleh untuk setiap kelasnya terdapat pada Gambar 7. Nilai yang lebih besar menunjukkan bahwa metode yang digunakan lebih baik. Nilai AUC yang didapatkan dari metode klasifikasi analisis diskriminan KTP ini berbeda untuk setiap kelas segmen yang terbentuk. Metode ini tidak cukup baik untuk memprediksi segmen yang terbentuk. Hal ini ditunjukkan oleh nilai AUC yang kurang dari 60% untuk setiap segmen. Segmen empat dapat diprediksi lebih baik dibandingkan dengan segmen yang lain. Segmen empat memiliki nilai AUC sebesar 58.15%. Segmen 2 memiliki nilai AUC terbesar kedua yakni 52.92% sedangkan segmen 1 dan 3 hanya memiliki nilai AUC sebesar 50.00% dan 51.95%.

4 Simpulan dan Saran

Segmentasi konsumen menggunakan pendekatan model RFM yang dilakukan menghasilkan 4 segmen yang terbagi menjadi segmen konsumen sangat loyal, loyal, tidak loyal, dan sangat tidak loyal. Banyaknya komponen KTP yang optimal dari



Gambar 7: Kurva ROC

analisis diskriminan kuadrat terkecil parsial adalah 8 komponen. Pemodelan klasifikasi ini tidak menghasilkan model yang cukup baik. Hal tersebut dapat dilihat dari plot dua skor komponen utama pertama yang tidak dapat membedakan setiap segmen. Nilai AUC yang diperoleh untuk menilai kinerja model juga tidak cukup baik yakni sebesar 50%, 52.92%, 51.95% dan 58.15% berturut-turut untuk segmen 1 sampai 4. Peubah penjelas yang berkontribusi tinggi terhadap model adalah X5, X9 dan X2 dengan nilai VIP lebih dari 1.

Saran untuk penelitian selanjutnya adalah membuat model klasifikasi lebih dari satu karena konsumen berasal dari satu Indonesia dengan asal provinsi yang berbeda-beda sehingga berpengaruh pada karakteristik konsumen yang berbeda pula. Selain itu, penelitian selanjutnya memanfaatkan peubah lain yang dapat menjelaskan loyalitas lebih baik sehingga dapat meningkatkan nilai AUC. Peubah-peubah tersebut misalnya jumlah anak, *rewards* yang ditukarkan oleh anggota, pendapatan konsumen, dan karakteristik provinsi yang lebih berhubungan dengan *supply* dan *demand* susu pertumbuhan untuk balita.

Daftar Pustaka

- Abdi, H. (2003). Partial Least Square Regression (PLS Regression), *Encyclopedia of Measurement and Statistics* **6(4)**: 792–795.
- Barker, M. and Rayens, W. (2003). Partial Least Squares for Discrimination, *Journal of Chemometrics* **17(3)**: 166–173.
- Berry, M. and Linoff, G. (2004). *Data Mining Techniques: For Marketing, Sales, and Customer Relationship Management*, John Willey Sons, Inc., Indiana(US).
- Galtier, O., Abbas, O., Le Dreau, Y., Rebufa, C., Kister, J., Artaud, J. and Dupuy, N. (2011). Comparison of PLS1-DA, PLS2-DA and SIMCA for Classification by Origin of Crude Petroleum Oils by MIR and Virgin Olive Oils by NIR for Different Spectra Regions, *Vibrational Spectroscopy* **55**: 132–140.
- Hosseini, S., Maleki, A. and Gholamian, M. (2010). Cluster analysis using data mining approach to develop CRM methodology to assess the customer loyalty, *Expert Systems with Applications* **37(7)**: 5259–5264.

- Hughes, A. (1994). *Strategic Database Marketing*, Probus Publishing Company, Chicago (US).
- Johnson, R. and Wichren, D. (2007). *Applied Multivariate Statistical Analysis*, Pearson Prentice Hall, New Jersey (US).
- Ngai, E., Xiu, L. and Chau, D. (2009). Application of data mining techniques in customer relationship management: A literature review and classification, *Expert systems with applications* **36(2)**: 2592–2602.
- Perez-Enciso, M. and Tenenhaus, M. (2003). Prediction of clinical outcome with microarray data: a partial least squares discriminant analysis (PLS-DA) approach, *Human Genetics* **112(5-6)**: 581–592.
- Tsiptsis, K. and Chorianopoulos, A. (2009). *Data Mining Techniques in CRM: Inside Customer Segmentation.*, John Willey Sons, Inc., West Sussex(UK).
- Wold, S., Sjostro, M. and Eriksson, L. (2001). PLS-Regression: A Basic Tool of Chemometrics, *Chemometrics and Intelligent Laboratory System* **58**: 109–130.