

---

# Penerapan *domain knowledge* pada model data nasabah perbankan dalam memprediksi *churn rate*

JBB  
10, 2

Gede Adi Aryanata, Linawati, Ida Bagus Alit Swamardika\*

*Universitas Udayana, Bali, Indonesia*

313

## ABSTRACT

*This study discusses the process of creating and optimizing data model which is used to predict the bank customer's churn rate based on several independent variables provided by the bank. This study used Logistic Regression as the main method to produce a data-to-data relationship model that describes the general classification of customers leaving banking services. It used the Backward Elimination method and the application of domain knowledge to optimize the performance of the resulting model. The last model produced is re-assessed using the Cumulative Accuracy Profile method. The results of this study have a good level of accuracy in predicting the churn rate. The model can predict 60% of people who will leave the bank by examining 29% of the overall data. The results of this research can be related to the level of business efficiency that can be done by banks in taking action on the customers with the highest probability level to leave banking services.*

## ABSTRAK

*Penelitian ini mendiskusikan tentang proses pembuatan dan pengoptimalan model data yang digunakan untuk memprediksi kemungkinan keluarnya nasabah dari suatu layanan perbankan berdasarkan beberapa variabel independen yang disediakan oleh pihak perbankan. Penelitian ini menggunakan Regresi Logistik sebagai metode utama untuk menghasilkan model hubungan antar data yang menggambarkan klasifikasi umum nasabah yang akan keluar dari layanan perbankan. Penelitian ini menggunakan metode Backward Elimination dan penerapan domain knowledge untuk mengoptimasi kinerja dari model yang dihasilkan. Model terakhir yang dihasilkan akan dinilai kembali dengan metode Cumulative Accuracy Profile. Hasil dari penelitian ini memiliki tingkat akurasi yang baik dalam memprediksi kemungkinan keluarnya nasabah dari layanan perbankan. Model dapat memprediksi 60% orang yang akan keluar dari bank dengan memeriksa 29% dari keseluruhan data. Hasil keluaran penelitian ini dapat dikaitkan dengan tingkat efisiensi usaha yang dapat dilakukan oleh pihak perbankan dalam mengambil tindakan pada nasabah dengan tingkat probabilitas tertinggi untuk keluar dari layanan perbankan.*

Keyword:

*Backward Elimination, Churn rate, Cumulative Accuracy Profile, Logistic Regression.*

## 1. PENDAHULUAN

Di era ini, sejumlah besar informasi dari berbagai bidang dikumpulkan dan disimpan untuk dapat dianalisis dan diekstraksi. Ini merupakan salah satu tugas paling menarik bagi perusahaan dan masyarakat untuk menemukan informasi dan pengetahuan baru dalam menunjang opini untuk proses pengambilan keputusan. Dengan munculnya pertanyaan-pertanyaan baru, berarti ini membutuhkan beberapa bidang ilmu untuk menyimpulkan solusi dari data yang didiskusikan. Bidang disiplin ilmu baru ini disebut *data science*.

*Received 8 Maret 2021*

*Revised 14 April 2021*

*Accepted 16 April 2021*

**JEL Classification:**

*E59, G21, G40*

**DOI:**

*10.14414/jbb.v10i2.2506*

**Journal of**

**Business and Banking**

ISSN 2088-7841

Volume 10 Number 2

November 2020 - April 2021

pp. 313-324

© STIE Perbanas Press

2021

Di dalam dunia bisnis, mencegah keluarnya pengguna dari layanan bisnis terkait merupakan tantangan para pelaku bisnis untuk dapat mempertahankan bisnisnya. Persentase tingkat keluarnya pengguna dari layanan bisnis sering disebut dengan *churn rate*. Dalam dunia perbankan, para peneliti berlomba-lomba untuk mencari pendekatan dan solusi dalam memprediksi nasabah yang akan keluar dari layanan perbankan dengan bertujuan menurunkan *churn rate*.

Di BPR XYZ, 99 dari 560 nasabah aktif pada Desember 2019, telah keluar dari layanan BPR XYZ pada Maret 2020. BPR XYZ kehilangan 17.7% nasabahnya dalam 3 bulan. Masalah ini merupakan salah satu contoh pentingnya penelitian lebih lanjut tentang metode atau cara yang dapat dilakukan untuk memprediksi nasabah yang akan keluar dari perbankan dengan tujuan menurunkan *churn rate* di BPR XYZ.

Dalam penerapan proses pengolahan data, ada beberapa penelitian (Samal et al., 2008; Golestan & Hezarkhani, 2016; Housni et al., 2019; Simumba et al., 2018) yang menggunakan metode regresi dalam melakukan klasifikasi suatu data. Selain regresi, ada beberapa metode lainnya yang dianggap baik dalam memetakan kumpulan data analisis dalam beberapa kategori. Beberapa dari penelitian (Samal et al., 2008; Golestan & Hezarkhani, 2016; Byana et al., 2018) juga menyimpulkan metode regresi dan *support vector machine* dianggap belum cukup untuk dalam membentuk model yang tepat. Penelitian tersebut menerapkan *Backward Elimination* sebagai metode tambahan untuk meningkatkan kualitas model yang dihasilkan. Pada penelitian yang dilakukan Zufa (2017), penerapan Regresi Logistik menjadi salah satu metode dalam mempertimbangkan kesehatan suatu perbankan. Dari keseluruhan penelitian, belum ada penelitian yang menerapkan dan memasukan *domain knowledge* dari spesifik bisnis area atau informasi dari para pekerja perbankan atas solusi dan pengalaman mereka dalam melakukan analisis untuk meningkatkan keluaran model. Proses implementasi *domain knowledge* dalam proses pembentukan model konseptual telah diabaikan dalam beberapa penelitian sebagaimana yang disampaikan oleh Khatri (2006).

Pada penelitian ini, penulis membuat sebuah analisis pemecahan masalah dalam memprediksi *churn rate* atau tingkat nasabah yang akan keluar dari layanan perbankan dengan menerapkan disiplin ilmu data science. Peneliti juga menerapkan implementasi *domain knowledge* dari perbankan untuk meningkatkan keluaran model yang dihasilkan.

Penelitian ini menggunakan tiga metode Regresi Logistik, *Backward Elimination* dan *Cumulative Accuracy Profile*. Keluaran analisis merupakan nilai probabilitas nasabah BPR XYZ yang akan berhenti menggunakan layanan dari BPR XYZ. Analisis akan menggunakan data yang diberikan oleh BPR XYZ dan hasil penelitian ini diharapkan dapat menjadi pacuan dan model data yang dapat digunakan dalam pengambilan keputusan di BPR XYZ.

## 2. RERANGKA TEORITIS DAN HIPOTESIS

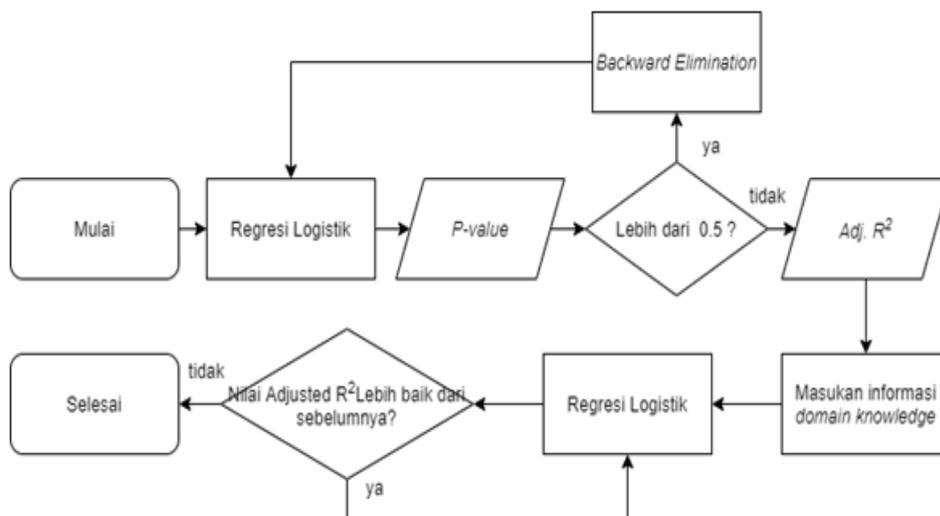
Penelitian ini menggunakan 3 metode utama dalam melakukan analisis. Analisis yang dihasilkan merupakan implementasi dari metode Regresi Logistik, *Backward Elimination* dan *Cumulative Accuracy Profile*. Kombinasi penerapan metode Regresi Logistik dan *Backward Elimination* serta

menerapkan *domain knowledge* dapat menghasilkan model data yang baik dalam memprediksi *churn rate* sesuai referensi penelitian Samal et al., 2008; Golestan & Hezarkhani, 2016; Byana et al., 2018 dan Khatri, 2006. Selain itu, *Cumulative Accuracy Profile* berperan dalam mengukur seberapa baik model yang dihasilkan dengan menurutkan dan memilih nasabah yang akan keluar dari layanan perbankan berdasarkan nilai probabilitas tertinggi yang dihasilkan. Proses analisis secara umum dapat ditampilkan pada Gambar 1.

**Regresi Logistik**

Regresi Logistik adalah metode yang tepat dalam menyelesaikan masalah yang memerlukan solusi 1 sampai 0 atau membagi sekumpulan data menjadi dua jenis kumpulan data. Metode ini diterapkan untuk menghasilkan model dari hubungan antarvariabel. Metode ini menghasilkan nilai probabilitas dengan mengabungkan beberapa variabel independen sebagai hasil prediksi dari kejadian yang akan datang. Salah satu contoh penerapan metode Regresi Logistik adalah menyelesaikan permasalahan dalam perbankan seperti mempertimbangkan penerimaan kredit bank berdasarkan beberapa variabel independen misalnya umur, penghasilan, jenis kelamin, dan lainnya sebagaimana dijabarkan dengan variabel X1 dan X2 pada persamaan berikut.

Hasil keluaran Regresi Logistik sangat berpengaruh pada cara pandang terhadap suatu variabel independen dalam model. Sebagai contoh, nilai keluaran Regresi Logistik terhadap data tenor dengan mendefinisikan tanggal awal nasabah bergabung dengan bank, dengan bentuk dd-mm-yy akan sangat berbeda dengan mendefinisikan data tenor dengan jumlah bulan atau jumlah hari sejak nasabah telah bergabung. Hal ini menyimpulkan hasil keluaran dari implementasi Regresi Logistik berpengaruh dengan seberapa baik cara pandang pengguna dalam mendefinisikan variabel independen yang digunakan.



**Gambar 1**  
Alur proses analisis

***Backward Elimination***

Metode *Backward Elimination* digunakan dalam mengoptimasi model awal yang dihasilkan. Menurut penelitian Samosir et al., (2014), metode ini bekerja dengan menghilangkan variabel independen dengan p-value tertinggi dari model. *Backward Elimination* pada umumnya menggunakan nilai 0.5 sebagai nilai ambang batas dari p-value. Proses ini akan dilakukan secara berulang sampai nilai p-value terendah di model tidak lebih dari 0.5. Selain p-value, nilai *Adjusted R-square* juga digunakan dalam menentukan kualitas dari model yang dihasilkan. Semakin baik nilai *Adjusted R-square* yang dihasilkan model, semakin baik model akan merepresentasikan hubungan antara data tersebut.

Penerapan metode *Backward Elimination* di penelitian ini dianggap belum cukup baik dalam mengoptimasi model yang dihasilkan. Beberapa variabel independen yang dihilangkan pada tahap ini dapat digunakan kembali setelah variabel independen divariasikan dengan variabel independen lainnya. Keluaran model akhir dalam implementasi *Backward Elimination* dapat digunakan sebagai standar dalam menentukan variabel independen yang berperan penting dalam melakukan variasi variabel lainnya.

***Cumulative Accuracy Profile***

Metode ini digunakan dalam menilai model yang dihasilkan. Dalam penelitian ini, metode ini akan menunjukkan seberapa baik model yang dihasilkan dalam memprediksi *churn rate*. CAP membandingkan antara keluaran prediksi dari model yang dihasilkan dengan model berdasarkan pengambilan acak dan jumlah dari data yang digunakan dalam setiap percobaan. Nilai keluaran dari model yang dihasilkan akan diurutkan berdasarkan nilai probabilitas tertinggi hingga terendah.

**3. METODE PENELITIAN****Pengumpulan Data**

Berdasarkan kebutuhan, maka data yang dikumpulkan untuk mendukung penelitian ini adalah penelitian-penelitian yang telah ada sebelumnya yang terkait dengan penelitian ini yaitu penelitian tentang *data science*, statistik, dan pengetahuan khusus seputar usaha kecil dan menengah, khususnya di bidang perbankan. Selain data penelitian yang dilakukan sebelumnya, data tentang profil nasabah dari usaha perbankan yang digunakan sebagai sumber data analisis juga diperlukan dalam penelitian ini.

Data nasabah diperoleh dari BPR XYZ dengan kondisi data identitas nasabah telah dihapus. Penelitian ini menggunakan 560 data nasabah BPR pada Desember 2019. Dari 560 nasabah tersebut, terdapat 99 nasabah keluar dari layanan perbankan pada Maret 2020. Peneliti juga melakukan proses wawancara pada pegawai BPR XYZ dan beberapa pegawai perbankan nasional lainnya untuk mendapatkan informasi seputar penerapan *domain knowledge* dalam model yang dihasilkan.

**Variabel Penelitian**

Berdasarkan kebutuhan, maka data yang dikumpulkan untuk mendukung penelitian ini adalah penelitian-penelitian yang telah ada sebelumnya yang terkait dengan penelitian ini yaitu penelitian tentang *data science*, statistik dan pengetahuan khusus seputar usaha kecil dan

---

menengah, khususnya di bidang perbankan. Selain data data penelitian yang dilakukan sebelumnya, data tentang profil nasabah dari usaha perbankan yang digunakan sebagai sumber data analisis juga diperlukan dalam penelitian ini.

### **Pengujian Model Analisis**

Pada tahap ini akan dilakukan pengujian terhadap sistem yang telah berhasil dibuat. Pengujian yang akan dilakukan adalah pengujian terhadap akurasi ratio model hasil prediksi dari hasil analisis dibandingkan model yang dihasilkan dengan pengambilan acak. Pengujian model akan mengacu pada nilai *Accuracy Ratio* (AR) dimana AR akan menentukan seberapa baik model dapat memprediksi keluarnya nasabah berdasarkan nilai probabilitas yang dihasilkan model. AR di definisikan sebagai rasio atau perbandingan antara daerah kurva hasil analisis dengan daerah kurva nilai sempurna.

## **4. ANALISIS DATA DAN PEMBAHASAN**

### **Analisis Model tahap 1**

Model pertama yang dihasilkan dengan implementasi Regresi Logistik menggunakan semua variabel independen yang diterima dari BPR sebagaimana yang ditampilkan pada Tabel 1. Variabel independen tersebut adalah umur, saldo tabungan terakhir, total produk yang digunakan, risiko kredit skor, estimasi pendapatan, transaksi terakhir, tenor, lokasi, dan jenis kelamin. Model pertama ini menghasilkan nilai *Adjusted R-square* 0.119119 dengan 82.3% nilai prediksi benar. Nilai prediksi benar ini mengklasifikasi nilai kebenaran menggunakan nilai 0.5 sebagai batas pembulatan. Nilai ini bukan pertimbangan utama dalam menentukan baik tidaknya model yang dihasilkan, akan tetapi dapat digunakan sebagai opini pendukung dalam penentuan model.

### **Analisis Model Tahap 2**

Analisis Model tahap 2 adalah analisis pada model dengan menerapkan metode Regresi Logistik dan *Backward Elimination*. Metode *Backward Elimination* bekerja dengan menghilangkan variabel independen dengan nilai p-value tertinggi dalam model. Dalam Tabel 1, risiko kredit skor memiliki nilai p-value tertinggi. Variabel independen ini akan dihapus dalam model dan metode Regresi Logistik akan diterapkan kembali tanpa variabel resiko kredit skor. Proses ini akan dilakukan secara berulang dan terhenti jika semua nilai independen variabel lebih kecil dari 0.5 dan atau model menghasilkan nilai *Adjusted R-square* tertinggi. Model terakhir yang dihasilkan dengan mengimplementasi metode *Backward Elimination* dijelaskan pada Tabel 2. Model ini menghasilkan nilai *Adjusted R-square* yang lebih baik dari model sebelumnya dengan nilai sebesar 0.132059 dengan 82.5% nilai prediksi benar. Hasil dari penerapan metode ini selaras dengan yang dengan hasil penelitian Darabi dan Hezarkhani (2016), Samal dan Ficarek (2008), Byana dan Abusa (2018), Sulaehani (2016) serta Bode A (2017). Dengan mengimplementasi metode ini, model telah menghilangkan 6 variabel independen. Variabel independen yang dihilangkan adalah total produk yang digunakan, resiko kredit skor, estimasi penghasilan, jenis kelamin dan salah satu lokasi BPR yang terdaftar.

**Tabel 1**  
**Model Tahap Regresi Logistik**

Var. Independen	Koefisien	P-Value
Constant	-3.64749	0.0009
Umur	0,0758106	0,0000000000597
Saldo Terakhir	0,000000109	0,0056
Jumlah Produk	-0,188670	0,3823
Tabungan di Bank	0,127670	0,6460
Resiko Kredit Skor	-0,00134952	0,2715
Jumlah bulan trans	0,211258	0,0000294
Estimasi Pemasuk	-0,000000283	0,4394
Tenor Edited Peny	-0,0784805	0,0608
Cabang B	-0,269289	0,3830
Cabang A	0,321060	0,3287
Pria	-0,288538	0,2440

Sumber : Hasil Penelitian (2020)

**Tabel 2**  
**Model Tahap Backward Elimination**

Var. Independen	Koefisien	P-Value
Constant	-5,2133	0,000000000000000798
Umur	0,0756615	0,000000000028
Saldo Terakhir	0,000000144	0,00000295
Jumlah bulan trans	0,203006	0,00000438
Tenor Edited Peny	-0,0734165	0,0744

Sumber : Hasil Penelitian (2020)

Kedua model yang dihasilkan pada Tabel 1 dan Tabel 2 memiliki kesamaan pada variabel independen umur yang memiliki nilai *p-value* terkecil dibandingkan variabel independen lainnya dalam masing-masing model tersebut. Hal ini merupakan informasi baru yang menyatakan variabel independen umur menjadi variabel independen yang memiliki pengaruh terbesar dalam model.

Proses implementasi informasi *domain knowledge* yang didapatkan dari BPR dan para pekerja profesional di area perbankan akan meningkatkan nilai *Adjusted R-square* dari model yang dibentuk. Beberapa informasi yang diterapkan dalam model ini adalah adanya korelasi yang terjadi antara umur dan beberapa variabel independen lainnya seperti tenor, estimasi pendapatan, kresiko kredit skor dan saldo tabungan untuk menentukan tingkat *churn rate* dari suatu nasabah. Dari Tabel 1 dan Tabel 2 dapat disimpulkan penelitian menghasilkan variabel independen umur memiliki nilai *p-value* terendah dalam model dan menjadi variabel yang memiliki tingkat pengaruh tertinggi dalam model. Kedua informasi dari pendapat para pelaku bisnis dan hasil analisis ditahap sebelumnya memiliki pandangan yang sama atas pengaruh variabel umur terhadap data.

Dari informasi tersebut, peneliti menggunakan kembali beberapa variabel independen yang telah dihapus dalam tahap implementasi *Backward Elimination*. Penggunaan variabel independen tersebut dilakukan bersamaan dengan melakukan variasi variabel independen tersebut dengan variabel independen umur. Hasil variasi bentuk variabel ini menghasilkan nilai *Adjusted R-square* dan nilai prediksi benar jauh lebih baik.

Beberapa variasi data yang dilakukan adalah melakukan nilai kuadrat pada saldo tabungan, melakukan pembagian pada tenor dan umur, melakukan logaritma dari hasil pembagian estimasi pemasukan dengan umur, melakukan logaritma pada hasil pembagian resiko kredit skor dengan umur serta melakukan logaritma dari hasil kuadrat nilai total produk yang dibagi dengan umur.

### Analisis Model Tahap 3

Analisis Model Tahap 3 adalah analisis model yang menerapkan metode Regresi Logistik, *Backward Elimination* dan *Domain knowledge*. Model terakhir yang dihasilkan dapat dilihat pada Tabel 3. Model ini dapat menggambarkan kumpulan hubungan data dengan baik dan menghasilkan nilai *Adjusted R-square* sebesar 0.155533 dengan 84.1% nilai prediksi benar. Nilai prediksi benar sebesar 84.1% diambil dengan menggunakan ambang batas nilai 0.5 untuk menentukan kategori prediksi benar atau prediksi salah. Table 4 menjelaskan secara ringkas ditampilkan nilai total positif palsu sebesar 75 nasabah, negative palsu sebesar 14 nasabah, total kasus positif sebesar 24 nasabah dan total kasus negatif sebesar 447 nasabah. Nilai yang dihasilkan sesuai dengan pernyataan pada penelitian sebelumnya tentang pentingnya penerapan *domain knowledge* dalam membentuk model dari beberapa variabel data.

**Tabel 3**  
**Model Tahap *Domain knowledge***

Var. Independen	Koefisien	P-Value
Constant	3,54726	0,0749
Jumlah bulan trans	0,224841	0,0000121
Cabang A	0,576220	0,0351
Kuadrat Saldo	0,00000	0,0027
Tenor Umur	-3,16286	0,0436
Log Pemasukan U	-0,529500	0,0390
Log Resiko Umur	-3,75736	0,0018
Log Kuadrat Jumla	0,735025	0,0041

Sumber : Hasil Penelitian (2020)

**Tabel 4**  
**Hasil Prediksi Model Tahap 3**

	Aktual Bertahan	Aktual Keluar
Diprediksi Bertahan	447	14
Diprediksi Keluar	75	24

Sumber : Hasil Penelitian (2020)

Selain melakukan variasi pada data, penelitian ini juga memeriksa *correlation matrix* dalam setiap variabel independen yang digunakan. *Correlation matrix* akan menjelaskan jika ada tingkat korelasi yang tinggi antar variabel independen. Semakin tinggi korelasi antara variabel, semakin buruk model yang dihasilkan. Buruknya model yang dihasilkan akibat tingginya *correlation matrix* dikarenakan oleh adanya penilaian ganda pada beberapa variabel independen dalam model. Tabel 5 memperlihatkan variabel independen Log Kuadrat Jumlah Umur dan Log Risiko Umur memiliki nilai tertinggi dalam *correlation matrix* dengan nilai 0.5081. Korelasi antar variabel dengan nilai lebih dari 0.7 dianggap memiliki tingkat korelasi yang kuat. Jika nilai berada di antara 0.5 dan 0.7 akan dianggap memiliki korelasi yang normal. Jika nilai berada di bawah 0.5 akan dianggap tidak memiliki korelasi.

Dalam tahap ini metode Cumulative Accuracy Profile (CAP) digunakan untuk melakukan pengujian terhadap model yang dihasilkan. Nilai probabilitas prediksi *churn* pada tahap ini diurutkan dari nilai tertinggi ke nilai terendah. Nilai probabilitas tertinggi menyatakan nasabah memiliki kemungkinan tertinggi untuk keluar dari layanan perbankan dan nasabah terkait diuji terlebih dahulu. Hal ini sejalan dengan proses bisnis perbankan yang akan menghubungi para nasabah yang memiliki kemungkinan keluar lebih tinggi dibanding yang tidak. Penerapan ini

**Tabel 5**  
*Correlation Matrix*

Var. 1	Var. 2	Correlation value
Jumlah bulan tra	Cabang A	0,0044
Jumlah bulan tra	Kuadrat Saldo	-0,0366
Jumlah bulan tra	Tenor Umur	0,1137
Jumlah bulan tra	Log Pemasuka	0,0626
Jumlah bulan tra	Log Resiko Um	0,0458
Jumlah bulan tra	Log Kuadrat Ju	-0,0981
Cabang A	Kuadrat Saldo	0,3425
Cabang A	Tenor Umur	0,0385
Cabang A	Log Pemasuka	0,0696
Cabang A	Log Resiko Um	0,0152
Cabang A	Log Kuadrat Ju	0,0138
Kuadrat Saldo	Tenor Umur	-0,0477
Kuadrat Saldo	Log Pemasuka	-0,0212
Kuadrat Saldo	Log Resiko Um	0,0160
Kuadrat Saldo	Log Kuadrat Ju	0,2232
Tenor Umur	Log Pemasuka	0,1742
Tenor Umur	Log Resiko Um	0,4118
Tenor Umur	Log Kuadrat Ju	-0,3733
Log Pemasuka	Log Resiko Um	0,2766
Log Pemasuka	Log Kuadrat Ju	-0,2530
Log Resiko Um	Log Kuadrat Ju	-0,5081

Sumber: Hasil Penelitian (2020)

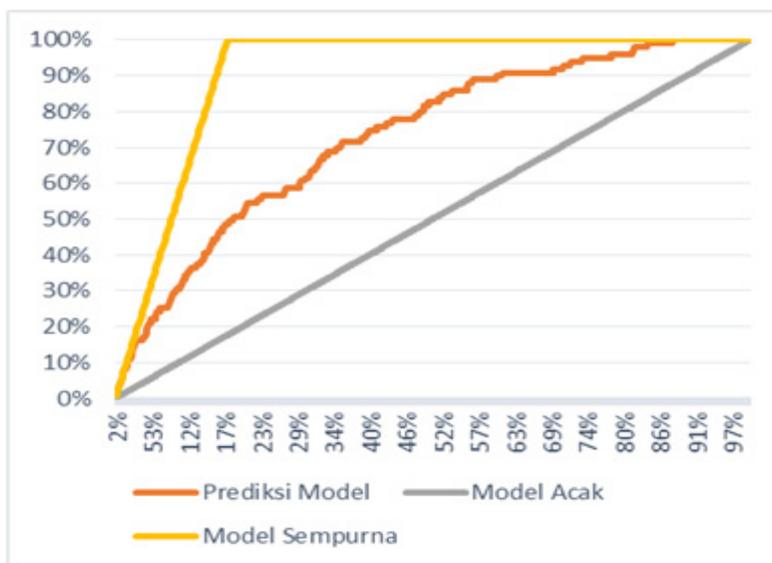
juga berkaitan dengan efisiensi usaha yang dilakukan pelaku perbankan dalam meminimalisir nasabah yang akan keluar dari layanan mereka. Pada Gambar 2 menjelaskan pengujian memperoleh hasil yang baik dalam memprediksi *churn rate* sebagai variabel dependen dalam model. CAP mengurutkan variabel dependen dari nilai probabilitas tertinggi yang dihasilkan model terakhir. Model dapat memprediksi 60% nasabah yang akan keluar dari layanan bank BPR dengan memeriksa 29% data nasabah. Nilai AR pada model tahap 3 yang mengacu pada gambar 2 adalah 0.57.

**Analisis Perbandingan 3 Model yang dihasilkan**

Penelitian ini menghasilkan 3 model pada 3 tahap yang berbeda. Dijelaskan di table 6, model pertama dengan metode Regresi Logistik menghasilkan nilai prediksi 82.3% dengan nilai *Adjusted R-square* 11.9%. Model kedua dengan menerapkan metode Regresi Logistik dan *Backward Elimination* menghasilkan nilai prediksi 82.5% dengan nilai *Adjusted R-square* 13,2%. Dan model ketiga yang menerapkan metode Regresi Logistik, *Backward Elimination* dan *domain knowledge* serta menerapkan variasi variabel menghasilkan nilai prediksi 84,1% dengan nilai *Adjusted R-square* 15.5%.

Semua model yang dihasilkan memiliki nilai *Adjusted R-square* yang sangat rendah. Nilai *Adjusted R-square* merupakan nilai yang melambangkan hubungan antar variabel independen satu dan variabel independen lainnya. Hal ini dapat disimpulkan, variabel independen yang diterapkan memiliki hubungan yang rendah.

Nilai 15.5% pada *Adjusted R-square* dapat menyimpulkan semua variabel independen yang digunakan pada model memiliki pengaruh sebesar 15.5% pada keluaran model atau variabel dependen, dimana variabel dependen merupakan variabel hasil prediksi. 84.5% variabel dependen dipengaruhi oleh variabel independen yang tidak digunakan pada penelitian ini.



**Gambar 2**  
**Grafik CAP Model Tahap 3**  
Sumber: Hasil Penelitian (2020)

**Tabel 6**  
**Perbandingan Model yang Dihasilkan**

Model	Metode	Nilai Prediksi	Adj R-square	Accuracy Ratio
Model 1	RL	82.3%	11.9%	58%
Model 2	RL BE	82.5%	13.2%	56%
Model 3	RL BE DK	84.1%	15.5%	57%

Sumber : Hasil Penelitian (2020)

Dalam wawancara dengan salah satu pegawai bank nasional dengan jabatan *senior customer service officer* didapatkan beberapa informasi pembandingan variabel independen yang dapat digunakan dalam menunjang model, di antaranya variabel tingkat pendidikan nasabah, variabel risk profile nasabah dan variabel tingkat kepuasan nasabah. Ketiga variabel ini merupakan aspek penting dalam menentukan *churn rate* nasabah perbankan.

Menurut pegawai bank tersebut, tingkat pendidikan berpengaruh pada respon nasabah saat melakukan verifikasi data atau memehuni dokumen persyaratan yang lebih kompleks. Sebagai contoh, mereka yang berpendidikan dibawah SMA, sering kali menginginkan proses yang tidak berbelit untuk mendapatkan jasa perbankan yang lebih mudah. Hal ini sering kali berujung pada nasabah yang batal melakukan transaksi perbankan. Selain itu, variabel risk profile didapatkan dengan melakukan pendataan pada nasabah terkait. Dari pengalaman pegawai tersebut, mereka yang cenderung berani mengambil resiko dalam berinvestasi sering kali mencari layanan perbankan lainnya untuk mendapatkan keuntungan lebih tinggi di dibandingkan mencari layanan yang aman.

Variabel tingkat kepuasan nasabah di bahas di penelitian Ahj & Lee (2018) tentang *churn rate* pada telekomunikasi serta penelitian dari Sunjaya (2018) tentang *churn rate* di perbankan, yang menjelaskan bagaimana kualitas layanan mempengaruhi tingkat *churn rate* di masing masing industri. Kedua penelitian ini menyampaikan pentingnya variabel tingkat kepuasan nasabah dalam model data untuk memprediksi pelanggan yang akan keluar dari layanan suatu bisnis.

##### 5. SIMPULAN, IMPLIKASI, SARAN, DAN KETERBATASAN

Penelitian ini menggambarkan pendekatan model statistik untuk memprediksi nasabah yang akan keluar dari layanan suatu BPR. Lima variabel baru dihasilkan dengan menerapkan variasi beberapa variabel independen utama dengan informasi yang diperoleh dari institusi terkait, para profesional di bidang perbankan dan dari hasil analisis ditahap sebelumnya. Hal ini merupakan pendekatan yang baik untuk menentukan apa yang harus diterapkan pada model untuk mendapatkan hasil yang lebih baik dalam menggambarkan hubungan antar data dalam kumpulan data.

---

Metode Regresi Logistik yang dioptimasi dengan metode *Backward Elimination* dianggap belum cukup dalam menghasilkan model yang baik. Penerapan *domain knowledge* atau pengetahuan tentang bidang bisnis terkait kedalam model dapat meningkatkan hasil keluar model menjadi lebih baik dari sebelumnya. Proses penemuan *domain knowledge* dapat dilakukan diluar insitusi tempat dilakukannya penelitian. Hal ini disarankan untuk dilakukan untuk mendapatkan pembandingan proses bisnis yang dilakukan oleh institusi satu dengan institusi lainnya.

**JBB**  
**10, 2**

Variabel independen pada dataset profil nasabah BPR XYZ dinilai belum memiliki hubungan yang cukup baik. Hal ini dapat dilihat dari rendahnya nilai *Adjusted R-square* yang dihasilkan model. Penerapan variabel pendidikan nasabah, variabel *risk profile* nasabah dan variabel tingkat kepuasan nasabah disarankan untuk mendapatkan keluaran model yang lebih baik. Model yang dihasilkan dapat menjadi acuan BPR XYZ dalam mengambil keputusan dalam memilih nasabah yang harus didekati untuk terindar dari keluar dari layanan BPR XYZ.

---

**323**

#### DAFTAR RUJUKAN

- Bode, A., 2017, K-Nearest Neighbor Dengan Feature Selection Menggunakan Backward Elimination Untuk Prediksi Harga Komoditi Kopi Arabika, *ILKOM Jurnal Ilmiah*, 9(2), 188–195. <https://doi.org/10.33096/ilkom.v9i2.139.188-195>
- Byana, A., & Abusa, F.N., 2018, Backward Elimination Untuk Meningkatkan Akurasi Kejadian Stunting Dengan Analisis Algoritma Support Vector Machine, *Dinamika Kesehatan*, Vol 9 No. 2 Desember 2018.
- Darabi-Golestan, F., & Hezarkhani, A., 2016, High precision analysis modeling by Backward Elimination with attitude on interaction effects on Au (Ag)-polymetallic mineralization of Glojeh, Iran. *Journal of African Earth Sciences*, 124(April 2018), 505–516. <https://doi.org/10.1016/j.jafrearsci.2016.09.030>
- Housni, M., Namir, A., Talbi, M., & Chafiq, N., 2019, Applying Data Analytics and Cumulative Accuracy Profile (CAP) Approach in Real-Time Maintenance of Instructional Design Models, *Advances in Intelligent Systems and Computing* (Vol. 756), Springer International Publishing, [https://doi.org/10.1007/978-3-319-91337-7\\_2](https://doi.org/10.1007/978-3-319-91337-7_2)
- Khatri, V., Vessey, I., Ramesh, V., Clay, P., & Park, S.-J., 2006, Understanding Conceptual Schemas: Exploring the Role of Application and IS Domain knowledge, *Information Systems Research* (17:1), pp. 81-99, <https://doi.org/10.1287/isre.1060.0081>

- Samosir, N., Siagian, P. and Bangun, P., 2014. Analisa Metode Backward dan Metode Forward untuk Menentukan Persamaan Regresi Linier Berganda (Kasus Jumlah Kecelakaan Lalu Lintas di Kotamadya). *Saintia Matematika*, 2(4), pp.345-360.
- Sunjaya, M.I., 2018, Tingkat Churn Tabungan Pada Industri Perbankan, *Jurnal Bisnis dan Manajemen*, Volume 2, Nomor 2, Hal. 157-169 <http://dx.doi.org/10.25139/ekt.v2i2.1101>.
- Samal, A. R., Mohanty, M. K., & Fifarek, R. H, 2008. *Backward Elimination* procedure for a predictive model of gold concentration, *Journal of Geochemical Exploration*, 97(2-3), 69-82. <https://doi.org/10.1016/j.gexplo.2007.11.004>.
- Simumba, N., Okami, S., Kodaka, A., & Kohtake, N, 2018, Alternative scoring factors using non-financial data for credit decisions in agricultural microfinance, *4th IEEE International Symposium on Systems Engineering, ISSE 2018 - Proceedings*, 1-8. <https://doi.org/10.1109/SysEng.2018.8544442>
- Sulaehani. R, 2016, Prediksi Keputusan Klien Telemarketing untuk Deposito Pada Bank Menggunakan Algoritma Naive Bayes Berbasis Backward Elimination, *ILKOM Jurnal Ilmiah*, Volume 8 Nomor 3 (Desember 2016). <https://doi.org/10.33096/ilkom.v8i3.83.182-189>.
- Zufa, F., Nugroho, S. and Simanihuruk, M., 2017. Perbandingan Analisis Diskriminan dan Analisis Regresi Logistik Ordinal dalam Prediksi Klasifikasi Kondisi Kesehatan Bank. *Jurnal Matematika*, 7(2), pp.92-106.

**\*Koresponden Penulis**

Penulis dapat dikontak pada e-mail: [gusalit@unud.ac.id](mailto:gusalit@unud.ac.id)