Evaluasi Telemarketing Kartu Kredit Bank Menggunakan Algoritma Genetika untuk Seleksi Fitur dan Naive Bayes

(Evaluation of Bank Credit Card Telemarketing Using Genetic Algorithms for Feature Selection and Naive Bayes)

Ekka Pujo Ariesanto Akhmad

Program Studi Ketatalaksanaan Pelayaran Niaga,
Program Diploma Pelayaran, Universitas Hang Tuah Surabaya

Abstrak: Bagian pemasaran bank sudah menampung data dari nasabah atau pelanggan bank dengan cara memasarkan atau mensosialisasikan kartu kredit lewat telepon (telemarketing). Evaluasi telemarketing kartu kredit yang sudah dilakukan bank masih kurang membawa hasil dan berdaya guna. Salah satu cara yang tepat untuk evaluasi laporan telemarketing kartu kredit bank adalah menggunakan teknik data mining. Tujuan penggunaan data mining untuk mengetahui kecenderungan dan pola nasabah yang berpeluang untuk berlangganan kartu kredit yang ditawarkan bank. Metode penelitian menggunakan Cross Industry Standard Process for Data Mining (CRISP-DM) dengan Algoritma Genetika untuk Seleksi Fitur (GAFS) dan Naive Bayes (NB). Hasil penelitian menunjukkan jumlah atribut pada dataset telemarketing kartu kredit bank sejumlah 15 atribut terdiri dari 14 atribut biasa dan 1 atribut spesial. Dataset telemarketing bank mengandung data berdimensi tinggi, sehingga diterapkan metode GAFS. Setelah menerapkan metode GAFS diperoleh 7 atribut optimal terdiri dari 6 atribut biasa dan 1 atribut spesial. Enam atribut biasa meliputi pekerjaan, balance, rumah, pinjaman, durasi, poutcome. Sedangkan atribut spesial adalah target. Hasil penelitian menunjukkan algoritma NB mempunyai nilai akurasi *86,71*%. Algoritma GAFS dan NB meningkatkan nilai akurasi menjadi *90,27*% untuk prediksi nasabah bank yang mengambil kartu kredit.
Kata kunci: telemarketing, CRISP data mining, algoritma genetika untuk seleksi fitur, data berdimensi tinggi, naive bayes

Abstract: *The bank marketing department has collected data from bank customers or customers by marketing or socializing credit cards by telephone (telemarketing). Evaluation of credit card telemarketing that has been done by banks still lacks results and is efficient. One appropriate way to evaluate bank credit card telemarketing reports is to use data mining techniques. The purpose of using data mining is to determine the trends and patterns of customers who have the opportunity to subscribe to credit cards offered by banks. The research method uses Cross Industry Standard Process for Data Mining (CRISP-DM) with Genetic Algorithms for Feature Selection (GAFS) and Naive Bayes (NB). The results showed the number of attributes in the bank credit card telemarketing dataset were 15 attributes consisting of 14 ordinary attributes and 1 special attribute. The bank telemarketing dataset contains high-dimensional data, so the GAFS method is applied. After applying the GAFS method, it is obtained that 7 optimal attributes consist of 6 ordinary attributes and 1 special attribute. Six ordinary attributes include pekerjaan, balance, rumah, pinjaman, durasi, poutcome. While special attributes are target. The results showed the NB algorithm has an accuracy value of 86.71%. GAFS and NB algorithms increase the accuracy value to 90.27% for predictions of bank customers who take credit cards.*
Keywords*: telemarketing, CRISP data mining, genetic algorithm for feature selection, high-dimensional data, naive bayes*

Alamat korespondensi:
Ekka Pujo Ariesanto Akhmad, Program Diploma Pelayaran, Universitas Hang Tuah, Jalan A. R. Hakim 150, Surabaya. e-mail: jurnal.pdp@hangtuah.ac.id

PENDAHULUAN

Salah satu bank BUMN menyediakan beberapa macam kartu kredit baru yang sesuai dengan kebutuhan dan gaya hidup untuk kenyamanan belanja kebutuhan sehari-hari, perjalanan (travel), otomotif, dan olahraga (golf). Kartu kredit tersebut akan ditawarkan kepada nasabah atau pelanggan bank.

Bagian pemasaran bank sudah menampung data dari nasabah atau

http://dx.doi.org/10.30649/japk.v10i1.71

pelanggan bank dengan cara memasarkan atau mensosialisasikan kartu kredit lewat telepon (telemarketing). Bagian pemasaran bank melakukan pengumpulan dan analisis data lewat telepon, namun masih kurang membawa hasil dan berdaya guna.

Salah satu cara yang tepat untuk evaluasi laporan telemarketing kartu kredit bank adalah menggunakan teknik data mining. Ringkasan data nasabah menggunakan data mining adalah untuk mengetahui kecenderungan dan pola nasabah yang berpeluang untuk berlangganan kartu kredit yang ditawarkan bank.

Telemarketing

Telemarketing adalah cara promosi yang masih dianggap efektif selain memasang pariwara di media publik. Pengertian telemarketing, yakni memasarkan atau mensosialisasikan produk atau jasa menggunakan telepon (Novia, 2017).

Algoritma Genetika

Algoritma Genetika atau Genetic Algorithm (GA) menyelesaikan permasalahan dengan mengimplementasikan seleksi alam, sehingga hanya individu terbaik yang bisa bertahan dan digunakan kembali pada proses seleksi selanjutnya. Solusi permasalahan diselesaikan dengan memetakan (*encoding*) ke dalam gen-gen, sehingga membentuk kromosom. Satu kromosom mewakili satu solusi dari suatu permasalahan. Pada penelitian ini, representasi kromosom yang digunakan yaitu menggunakan biner.

Setelah itu terdapat proses reproduksi yang bertujuan untuk menghasilkan *offspring* yang beragam. Offspring beragam menunjukkan banyaknya variasi calon solusi yang diperoleh. Proses reproduksi terdiri dari proses *crossover* dan proses mutasi (Fadilah, Cholissodin, & Mahmudy, 2015). Proses crossover yang digunakan adalah menggunakan jenis *one-cut-point* dengan cara menyilangkan gen pada *point cut* tertentu dari dua individu yang dipilih secara acak. Sedangkan proses reproduksi selanjutnya adalah mutasi yang dilakukan dengan cara memilih salah satu poin gen secara acak kemudian mengubahnya menjadi nilai biner kebalikannya.

Proses selanjutnya adalah menghitung nilai *fitness.* Fitness digunakan untuk menentukan seberapa optimal calon solusi untuk menyelesaikan permasalahan. Semakin besar nilai fitness, maka calon solusi tersebut memiliki kemungkinan lebih besar dan baik untuk dijadikan solusi.

Algoritma Naive Bayes

Naive Bayes (NB) merupakan teknik sederhana yang sangat cocok saat dimensi dari input tinggi. Meskipun sederhana, Naive Bayes seringkali dapat mengungguli metode lain yang lebih canggih. Untuk membangun model Naive Bayes, perlu menghitung probabilitas setiap nilai atribut yang diberi atribut class (Bustami, 2013).

Persamaan dari teorema Bayes adalah

$$P(H|X) = \frac{P(X|H).P(H)}{P(X)}$$

Keterangan:

X : Data dengan class yang belum diketahui

H : Hipotesis data X merupakan suatu class spesifik

P(H|X): Probabilitas hipotesis berdasar kondisi (posteriori probability)

P(H) : Probabilitas hipotesis H (prior probability)

P(X|H): Probabilitas X berdasarkan kondisi pada hipotesis H

P(X) : Probabilitas X

Seleksi Fitur

Metode seleksi fitur diterapkan untuk menangani data berdimensi tinggi (Canedo and Noelia, 2015).

Data berdimensi tinggi mengakibatkan hasil akurasi klasifikasi rendah (Saez, et al., 2013).

http://dx.doi.org/10.30649/japk.v10i1.71

Tabel 1
Hasil Penelitian yang Relevan

Peneliti/ tahun	Judul	Masalah	Pendekatan/ Teori	Hasil
Aqham, Ahmad Asifuddin dan Kristoko Dwi Hartomo (2019)	Data Mining untuk Nasabah Bank Telemarketing Menggunakan Kombinasi Algoritm Naïve Bayes dan Algoritma Genetik	Bagaimana mengetahui kendala nasabah telemarketing yang ingin melakukan deposito pada bank	- Naïve Bayes (Bustami, 2013; Zaidi, d.k.k., 2013) - Feature Selection (Kusumadewi, 2003) - Algoritma Genetika (Zukhri, 2014; Trevino, 2006)	Penentuan Nasabah Bank yang akan Berdeposito dengan *feature selection*, dengan hasil nilai 86,71% menjadi 90,27%, Peningkatan akurasi yang dihasilkan tentunya belum sempurna, sehingga perlu adanya upaya-upaya lain dalam meningkatkan akurasi yang dihasilkan.
Oktanisa, Irvi dan Ahmad Afif Supianto (2018)	Perbandingan teknik klasifikasi dalam data mining untuk bank direct marketing	Bagaimana mencapai akurasi yang tinggi dalam proses klasifikasi berdasarkan ketepatan informasi tertentu yang diperoleh dari *customer* dan dianggap penting oleh pihak bank	- SVM (Shmilovici, 2009; Suthaharan, 2016) - AdaBoost (Hu, d.k.k, 2008) - Naive Bayes (Lewis, 1998; Genkin, 2007; Karim, 2013) - Constant (Flach, 2016) - K-Nearest Neighbor (Afandie, 2014; Ekaristio, 2015) - Tree (Quinlan, 1987; Klas & Schrefl, 1995) - Random Forest (Breiman, 1999; Pal, 2005) - Stochastic Gradient Descent (Mandt, 2017) - CN-2 Rule (Clark & Boswell, 1991)	Setelah dilakukan pengujian dengan menggunakan *tool*, didapat hasil klasifikasi dengan akurasi yang tinggi pada model *Tree*, *Naive Bayes*, *Constant* dan *SGD*. Disusul dengan model *random forest*, *KNN*, *CN-2 Rule Inducer*, *AdaBoost*, dan SVM. Kesembilan model yang diusulkan memiliki karakteristik masing-masing model berdasarkan spesifikasi datanya. Berdasarkan visualisasi yang dihasilkan dari pengujian menggunakan *tool* Orange, terlihat bahwa *Stochastic Gradient Descent* (SGD) memiliki visualisasi yang terbaik untuk mencapai target tujuan klasifikasi untuk *bank direct marketing* yaitu diterima (*yes*) dan ditolak (*no*).
Suhaelani, Ruhmi (2016)	Prediksi keputusan klien *telemarketing* untuk deposito pada bank menggunakan algoritma *naive bayes* berbasis *backward elimination*	Sulitnya mengetahui keputusan klien *Telemarketing* untuk melakukan deposito pada bank, menyebabkan bank selalu menghadapi ancaman krisis keuangan.	- Naive Bayes (Bustami, 2013; Aggarwal C. & Philip S. Yu, 2008; Valle, dkk, 2012) - Backward Elimination (Chana K. Y., dkk., 2011)	Dalam pemilihan atribut yang berpengaruh dilihat dari nilai *F* terkecil dan nilai *signifikan*, jika nilai *F* terkecil dari suatu atribut memiliki nilai *signifikan* dari nilai alpha <=0,05 maka atribut tersebut berpengaruh, dari 14 atribut yang digunakan yaitu *age*, *duration*, *campaign*, *previous*, *job*, *marital*, *education*, *default*, *housing*, *loan*, *contact*, *month*, *day_of_week* dan *poutcome* hanya 4 atribut yang berpengaruh dari hasil evaluasi yaitu *duration*, *campaign*, *contact* dan *poutcome*. Hasil evaluasi *confusion matrix* dengan menggunakan algoritma *Naive Bayes* berbasis *Backward Elimination* di dapatkan akurasi yang tinggi yaitu 90,69% dibanding dengan hanya menggunakan algoritma *Naive Bayes* saja yaitu 89,08%.
Maghfirah, Teguh Bharata Adji, dan Noor Akhmad Setiawan (2015)	Menggunakan *data mining* untuk segmentasi customer pada Bank untuk meningkatkan *Customer Relationship Management* (CRM) dengan metode klasifikasi (algoritma *J-48*, *Zero-r* dan *Naive Bayes*)	Bank tidak dapat menarik semua nasabah di pasar tersebut. Hal ini disebabkan : nasabah terlalu banyak, nasabah tersebar luas, nasabah mempunyai kebutuhan dan kebiasaan yang bervariasi.	Naive Bayes, Rules Zero-R, Trees-J48 (K. Wisaeng, 2013)	Untuk dataset yang digunakan yaitu bank dataset, metode classification dengan algoritma J-48 memberikan hasil terbaik dengan error rate terkecil, yaitu 5.8769%. Sedangkan jika dilihat dari segi efisiensi waktu dan hasil klasifikasi, algoritma Zero-R memberikan hasil terbaik (0,03 detik), walaupun tidak terlalu signifikan dibandingkan dengan Naive Bayes. Kemudian instances kelas Yes yang paling banyak mengalami kesalahan klasifikasi adalah ketika menggunakan Zero-R, sementara untuk instances kelas No yang paling banyak mengalami kesalahan klasifikasi adalah ketika menggunakan Naive Bayes.

Sumber: Aqham (2019), Oktanisa (2018), Suhaelani (2016), Maghfirah (2015)

http://dx.doi.org/10.30649/japk.v10i1.71

Algoritma genetika untuk seleksi fitur (*Genetics Algorithm for Feature Selection*) digunakan untuk meningkatkan hasil klasifikasi yang rendah, supaya lebih baik.

Cross Industry Standard Process for Data Mining (CRISP-DM)

Metodologi CRISP-DM mempunyai enam tahapan, yaitu Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, dan Deployment (Suntoro, 2019).

Business understanding berisi tentang menentukan tujuan bisnis, menilai situasi saat ini, dan menetapkan tujuan dilakukan data mining.

Data understanding merupakan kegiatan persiapan, mengevaluasi persyaratan data, dan termasuk pengumpulan data.

Data preparation dilakukan setelah data dikumpulkan, data-data tersebut perlu diidentifikasi, dipilih, dibersihkan, kemudian dibangun ke dalam bentuk/format yang diinginkan.

Modeling adalah aplikasi dari algoritma untuk mencari, mengidentifikasi, dan menampilkan pola.

Evaluation digunakan untuk membantu pengukuran evaluasi pada model.

Deployment digunakan untuk melakukan otomatisasi model atau pengembangan aplikasi, terintregasi dengan sistem informasi manajemen atau operasional yang ada.

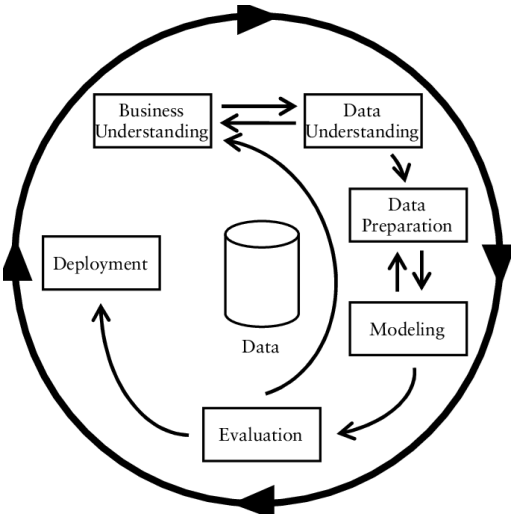
Penelitian yang relevan

Tabel 1 menunjukkan hasil penelitian terdahulu yang telah dilakukan peneliti-peneliti lain sebelumnya. Persamaan penelitian yang dilakukan dengan penelitian terdahulu adalah menggunakan algoritma Genetik dan Naive Bayes. Penelitian yang dilakukan memiliki perbedaan dengan penelitian terdahulu, yaitu memakai metode CRISP-DM untuk telemarketing kartu kredit bank.

METODE PENELITIAN

Penelitian ini menggunakan metodologi data mining CRISP-DM.

Gambar 1 menjelaskan enam tahapan CRISP-DM (Shearer, 2000).



Gambar 1. Tahapan CRISP-DM
Sumber: Shearer (2000)

HASIL DAN PEMBAHASAN

Penelitian ini mengerjakan tahapan CRISP-DM sebagai berikut.

Business Understanding

Bagian pemasaran bank telah melakukan pemasaran kartu kredit dan mengumpulkan data lewat panggilan telepon (telemarketing) kepada nasabah yang telah menjadi pelanggan bank, untuk mengetahui apakah nasabah bersedia untuk berlangganan kartu kredit atau tidak. Cara tersebut dirasa tidak efektif dan efisien.

Salah satu cara yang efektif untuk evaluasi laporan telemarketing kartu kredit bank adalah menggunakan teknik data mining.

Data Understanding

Data nasabah bank sudah dikumpulkan oleh bagian pemasaran bank sebanyak 4.521 nasabah (Moro, et al., 2014). Data nasabah terdiri dari 15 atribut. Kelima belas atribut data nasabah adalah sebagai berikut.

a. Usia, berisi data usia nasabah, tipe data integer.

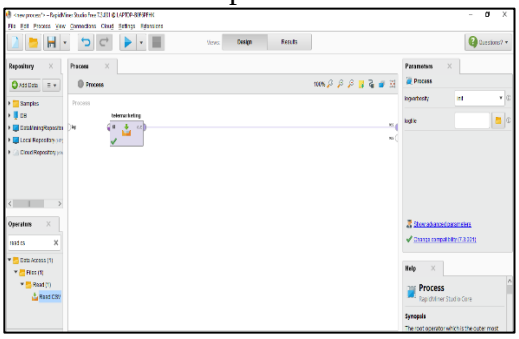
http://dx.doi.org/10.30649/japk.v10i1.71

b. Pekerjaan, berisi data pekerjaan dari nasabah, tipe data polynominal.
c. Status, berisi data status pernikahan nasabah, tipe data polynominal.
d. Pendidikan, berisi data pendidikan terakhir dari nasabah, tipe data polynominal.
e. Kartu Kredit, berisi data apakah nasabah mempunyai kartu kredit atau tidak, tipe data binominal.
f. Balance, berisi data jumlah uang yang disimpan di bank (dalam puluhan ribu rupiah), tipe data integer.
g. Rumah, berisi data apakah nasabah mempunyai pinjaman untuk kredit rumah atau tidak, tipe data binominal.
h. Pinjaman, berisi data apakah nasabah mempunyai pinjaman untuk kredit selain rumah atau tidak, tipe data binominal.
i. Media, berisi data media komunikasi yang digunakan untuk menawarkan produk kartu kredit antara bagian pemasaran dan nasabah, tipe data polynominal.
j. Bulan, berisi data periode bulan terakhir kali nasabah dihubungi oleh bagian pemasaran, tipe data polynominal.
k. Durasi, berisi data durasi waktu lamanya panggilan, tipe data integer.
l. Persuasi, berisi data jumlah kontak yang sudah dilakukan antara bagian pemasaran dan nasabah selama telemarketing kartu kredit berlangsung, tipe data integer.
m. Previous, berisi data jumlah kontak yang telah dilakukan antara bagian pemasaran dan nasabah sebelum diadakannya telemarketing kartu kredit, tipe data integer.
n. Poutcome, berisi data hasil dari telemarketing, tipe data polynominal.
o. Target, adalah kelas berisi data ya, jika nasabah akan berlangganan kartu kredit yang telah ditawarkan bagian pemasaran atau berisi data tidak, jika nasabah tidak akan berlangganan kartu kredit yang dipromosikan bagian pemasaran bank, tipe data binominal.

Data Preparation

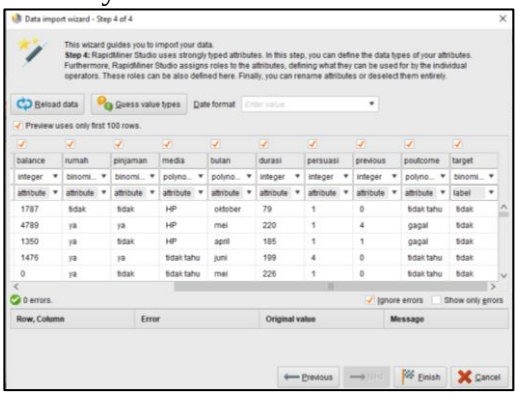
Perangkat lunak yang digunakan untuk pengolahan dan visualisasi data pada penelitian ini adalah RapidMiner 7.3. Sesuai dengan data understanding yang sudah diterangkan sebelumnya diketahui dataset telemarketing terdiri dari 4.521 baris dan 15 atribut.

Gambar 2 menjelaskan dataset telemarketing yang telah disimpan dalam format file csv dimasukkan ke dalam Process RapidMiner.



Gambar 2. Read CSV pada RapidMiner
Sumber: file process telemarketing RapidMiner

Gambar 3 menunjukkan data telemarketing diatur tipe data dan jenis atributnya.



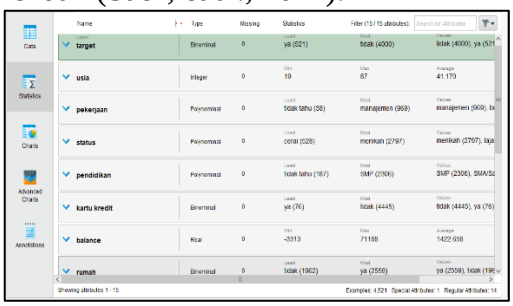
Gambar 3. Atur Tipe Data dan Jenis Atribut
Sumber: file process telemarketing RapidMiner

Gambar 4 menerangkan pada tab Statistics, melihat metadata dari dataset telemarketing bank tidak ada missing value.

Namun, kendala yang timbul pada dataset telemarketing bank adalah data mempunyai banyak atribut atau data berdimensi tinggi. Masalah dalam data mining khususnya klasifikasi adalah data berdimensi tinggi (Jian, C.,et al., 2016). Data berdimensi tinggi

http://dx.doi.org/10.30649/japk.v10i1.71

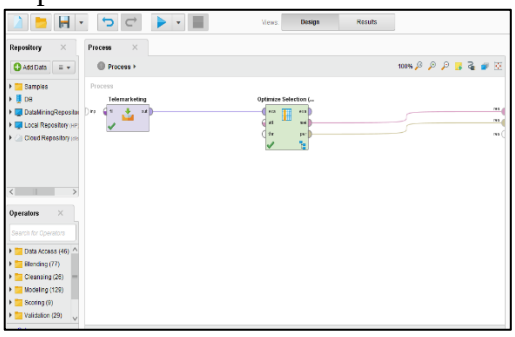
menyebabkan ukuran dataset menjadi lebih besar, jumlah atribut yang banyak, atau jumlah data sampel yang besar (Bennasar, et al., 2015). Data berdimensi tinggi menyebabkan performa algoritma klasifikasi menjadi rendah (Saez, et al., 2014).



Gambar 4. Metadata dataset telemarketing
Sumber: file process telemarketing RapidMiner

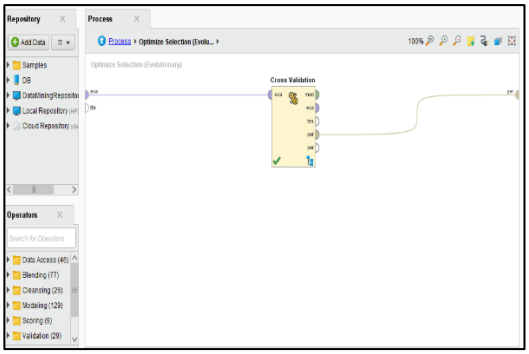
Data berdimensi tinggi dapat diselesaikan dengan seleksi fitur (Canedo and Noelia, 2015). Seleksi fitur berguna untuk mengurangi ukuran atribut yang besar pada dataset (Sebastiani, 2002). Salah satu metode seleksi fitur yang digunakan untuk mengurangi jumlah atribut adalah Genetic Algorithm based Feature Selection (GAFS) (Wahono dan Herman, 2014). Genetic Algorithm digunakan untuk mencari solusi atribut paling optimal. Penerapan GAFS pada Rapdiminer adalah sebagai berikut.

Gambar 5 menjelaskan operator Optimize Selection (Evolutionary) ditambahkan pada halaman view design RapidMiner.



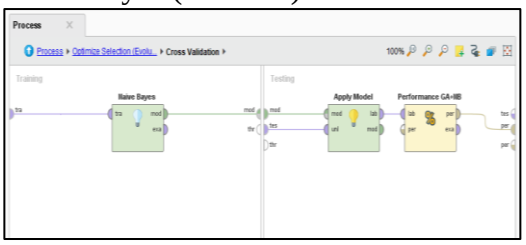
Gambar 5.Penambahan operator Optimize Selection (Evolutionary)
Sumber: file process telemarketing RapidMiner

Gambar 6 menjelaskan operator Optimize Selection (Evolutionary) diklik ganda untuk menambahkan Cross Validation.



Gambar 6. Penambahan operator Cross Validation
Sumber: file process telemarketing RapidMiner

Gambar 7 menerangkan operator Cross Validation diklik ganda untuk menambahkan tiga operator, yaitu Naive Bayes, Apply Model, dan Performance Genetic Algorithm dan Naive Bayes (GA+NB).



Gambar 7. Penambahan operator Naive Bayes, Apply Model, dan Performance
Sumber: file process telemarketing RapidMiner

Kemudian, tombol Run pada halaman Process diklik. Setelah tombol Run diklik akan terlihat AttributeWeights. Atribut yang dipilih berdasarkan metode GAFS adalah atribut dengan bobot sama dengan satu, yaitu atribut usia, pendidikan, kartu kredit, balance, rumah, durasi, dan poutcome. Sedangkan atribut dengan bobot sama dengan nol akan dibuang atau tidak terpilih.

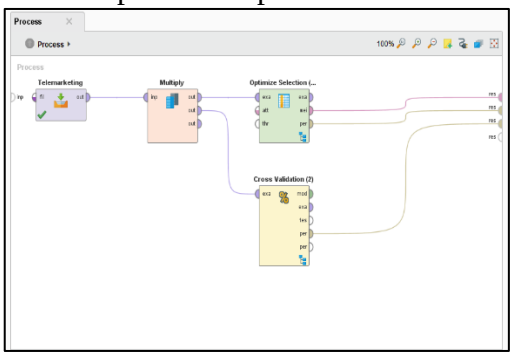
Modeling

Metode seleksi fitur GAFS digunakan pada tahapan data preparation. Pada tahapan modeling, algoritma Naive Bayes akan dibandingkan dengan algoritma GAFS+NB. Sehingga, tahapan modeling akan melanjutkan tahapan data preparation.

Gambar 8 menunjukkan operator Multiply ditambahkan pada

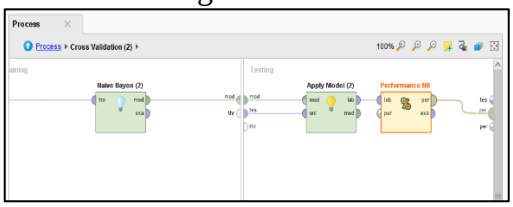
http://dx.doi.org/10.30649/japk.v10i1.71

halaman process RapidMiner. Operator Multiply digunakan untuk menyalin data pada port inputnya ke port output beberapa kali. Setelah itu, operator Cross Validation ditambahkan ke dalam halaman process RapidMiner.



Gambar 8. Penambahan operator Multiply dan Cross Validation (2)
Sumber: file process telemarketing RapidMiner

Gambar 9 menerangkan operator Cross Validation diklik ganda untuk menambahkan operator Naive Bayes. Bersamaan dengan penambahan operator Naive Bayes, ditambahkan pula operator Apply Model dan Performance. Nama operator Performance diganti Performance NB.



Gambar 9. Penambahan operator Naive Bayes (2), Apply Model (2), dan Performance NB
Sumber: file process telemarketing RapidMiner

Kemudian, tombol Run diklik pada halaman Process RapidMiner.

Evaluation

Bank menerapkan Genetic Algorithm for Feature Selection (GAFS) untuk mengatasi masalah data telemarketing dengan atribut yang banyak. Genetic Algorithm digunakan untuk mencari solusi atribut paling optimal. Dataset telemarketing kartu kredit bank akan dihitung akurasinya dengan membandingkan antara algoritma NB dan algoritma GAFS+NB. Metode yang terbaik akan digunakan untuk tahapan deployment.

Pada tahapan modeling dihasilkan convolution matrix seperti yang terlihat pada Tabel 2 untuk hasil algoritma NB dan Tabel 3 untuk hasil algoritma GAFS+NB.

Tabel 2
Convolution Matrix Metode NB

	true tidak	true ya
pred. tidak	3.738	273
pred. ya	262	248

Tabel 3
Convolution Matrix Metode GAFS+NB

	true tidak	true ya
pred. tidak	3.870	317
pred. ya	130	204

Setelah hasil convolution matrix diperoleh, maka langkah berikutnya adalah menghitung nilai akurasi masing-masing metode. Nilai akurasi untuk masing-masing metode adalah sebagai berikut.

$$\text{Akurasi NB} = \frac{3738 + 248}{3738 + 273 + 262 + 248} = 88{,}17\%$$

$$\text{Akurasi GAFS+NB} = \frac{3870 + 204}{3870 + 317 + 130 + 204} = 90{,}11\%$$

Berdasarkan penghitungan nilai akurasi di atas, maka dapat disimpulkan bahwa metode GAFS+NB lebih baik daripada metode NB.

Berdasarkan kesimpulan tersebut, bahwa data berdimensi tinggi menyebabkan performa (hasil akurasi) algoritma klasifikasi menjadi rendah (Saez, et al., 2014) dan salah satu metode untuk mengatasi data berdimensi tinggi adalah menerapkan metode seleksi fitur (Canedo and Noelia, 2015).

Deployment

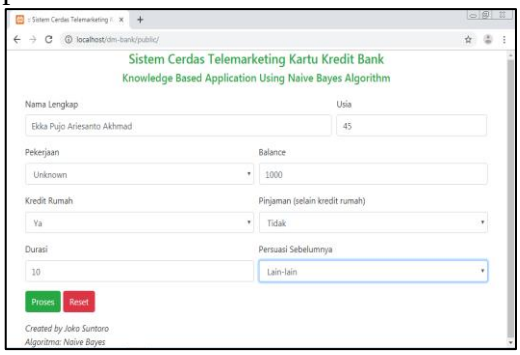
Pada tahapan evaluation didapatkan hasil akurasi metode GAFS+NB lebih baik daripada metode NB. Jumlah atribut pada dataset telemarketing kartu kredit bank sejumlah 15 atribut terdiri dari 14 atribut biasa dan 1 atribut spesial.

http://dx.doi.org/10.30649/japk.v10i1.71

Dataset telemarketing bank mengandung data berdimensi tinggi, sehingga diterapkan metode GAFS. Setelah menerapkan metode GAFS diperoleh 7 atribut optimal terdiri atas 6 atribut biasa dan 1 atribut spesial. Enam atribut biasa meliputi pekerjaan, balance, rumah, pinjaman, durasi, poutcome. Sedangkan atribut spesial adalah target.

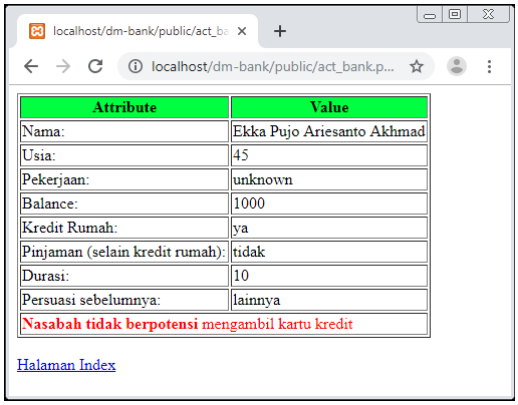
Sistem berbasis web akan dibangun pada tahapan deployment untuk mencari nasabah yang mungkin berlangganan kartu kredit bank. Parameter yang dipakai untuk pencarian nasabah yang mungkin berdasarkan tujuh atribut yang telah diseleksi menggunakan metode GAFS.

Evaluasi telemarketing kartu kredit bank berbasis web dapat dilihat pada Gambar 10.



Gambar 10.Halaman awal evaluasi telemarketing kartu kredit bank
Sumber: file index.php

Gambar 11 menunjukkan hasil proses evaluasi telemarketing kartu kredit bank.



Gambar 11. Halaman hasil proses evaluasi telemarketing kartu kredit bank
Sumber: file act_bank.php

KESIMPULAN DAN SARAN

Kesimpulan

Evaluasi telemarketing kartu kredit bank menggunakan Algoritma Genetika untuk Seleksi Fitur dan Naive Bayes memberikan hasil algoritma NB nilai akurasi 86,71% menjadi 90,27% dengan algoritma GAFS dan NB untuk prediksi nasabah bank yang mengambil kartu kredit.

Saran

Peningkatan akurasi dihasilkan penelitian ini belum sempurna, sehingga perlu mencoba metode seleksi fitur lain, misal CBS, PCA, dan lain-lain, serta algoritma lain, seperti Neural Network (NN), Particle Swarm Optimization (PSO), dan lain-lain untuk meningkatkan akurasi yang dihasilkan.

DAFTAR PUSTAKA

Afandie, M. N., Cholissodin, I., & Supianto, A. A. (2014). Implementasi metode k-nearest neighbor untuk pendukung keputusan pemilihan menu makanan sehat. *Repository Jurnal Mahasiswa PTIIK UB*, *3*(1), 1.

Aggarwal, C. & Philip S. Yu. (2008). Privacy-Preserving Data mining : Models and Algorithms, Springer.

Aqham, Ahmad Asifuddin dan Kristoko Dwi Hartomo. (2019). Data Mining untuk Nasabah Bank Telemarketing Menggunakan Kombinasi Algoritm Naïve Bayes dan Algoritma Genetik. InfoTekJar: *Jurnal Nasional Informatika dan Teknologi Jaringan* - Vol. 4 No. 1 edisi September.

Bennasar, M., Hicks, Y., & Setchi, R. (2015). Feature selection using Joint Mutual Information Maximisation. *Expert Systems with Applications*, 42(22), 8520-8532,https//doi.org/10.1016/j.eswa.2015.07.007

Bustami, B. (2013). "Penerapan Algoritma Naïve Bayes Untuk Mengklasifikasi Data Nasabah Asuransi," *TECHSI : Jurnal Penelitian Teknik Informatika*, vol. 3, no. 2, pp. 127 – 146.

Breiman, L. (1999). Random forests – random features, Technical Report 567, Statistics Departement, University of California, Berkeley.

Canedo, Veronica Bolon dan Noelia Sanchez Marono, A. A.-B. (2015). Feature selection for high-dimensional industrial data. Springer International Publishing Switzerland 2015.https://doi.org/10.1007/s13748-015-0080-y

Chana K. Y., Kwong. C. K, Dillona. T. S., and Tsimb Y. C. (2011). Reducing Overfitting in Manufacturing Process Modeling Using a Backward Elimination Based Genetic Programming, *Applied Soft Computing*.

Clark, P. & Boswell, R. (1991). Rule induction with CN2: Some recent improvements, In: Kodratoff Y.(eds) Machine Learning – EWSL-91, EWSL 1991, Lecture Notes in Computer Science (Lecture Notes in Artificial Intelligence), vol 482, Springer, Berlin, Heidelberg.

Ekaristio, I., Soebroto, A. A., & Supianto, A. A. (2015). Pengembangan sistem pendukung keputusan pemilihan bibit unggul sapi bali menggunakan metode k-nearest neighbor. *Journal of Environmental Engineering and Sustainable Technology, 02*(01), 49–57.

Fadilah, A. N., Cholissodin, I., dan Mahmudy, W. F. (2015). Implementasi Analytical Hierarchy Process (AHP) dan Algoritma Genetika untuk Rekomendasi dan Optimasi. *DORO: Repository Jurnal Mahasiswa PTIIK Universitas Brawijaya*, *5*(14).

Flach, P. A. (2016). Classifier calibration, In: C. Sammut, G.I., Webb (eds), Encyclopedia of machine learning and data mining, Springer, Boston, MA.

Genkin, A., Lewis, D. D., & Madigan, D. (2007) Large-scale Bayesian logistic regression for text categorization, *Technometrics*, vol.49, pp. 291-304.

Hu, W. & Maybank, S. (2008). AdaBoost-based algorithm for network instrusion detection, IEEE Transactions on Systems, Man, and Cybernetics - Part B: Cybernatics, vol.38, no.2.

Jian, C., Gao, J., & Ao, Y. (2016). A new sampling method for classifying imbalanced data based on Support Vector Machine ensemble. *Neurocomputing* 193, 115-122. https://doi-org/10.1016/j.neucom.2016.02.006

K. Wisaeng. (2013). "A Comparison of Different Classification Techniques for Bank Direct Marketing". *International Journal of Soft Computing and Engineering (IJSCE)*.

Karim, M., & Rahman, R. M. (2013). Decission tree and naïve bayes algorithm for classification and generation of actionable knowledge for direct marketing, *Journal of Software Engineering and Application*, vol.6,pp.196-206.

Klas, W., & Schrelf, M. (1995). Metaclasses and their application: Data model tailoring and database integration. Springer.

Kusumadewi, Sri. (2003). Artificial Intelligence (teknik dan aplikasinya), Yogyakarta: Graha Ilmu.

Lewis, D. D. (1998). Naïve (Bayes) at forty: The independence assumption in information retrieval, in European Conference on Machine Learning, pp. 4-15.

Maghfirah, d.k.k. (2015). Menggunakan *data mining* untuk Segmentasi Customer pada Bank untuk Meningkatkan *Customer Relationship Management* (CRM) dengan Metode Klasifikasi (Algoritma *J-48*, *Zero-r* dan *Naive Bayes*). *Prosiding SNST* ke-6, Fakultas Teknik Universitas Wahid Hasyim Semarang.

Mandt, S., Hoffman, M. D., & Blei, D. M. (2017). Stochastic gradient descent as approximate Bayesian inference, *Journal of Machine Learning Research*, 18, 1-35.

North, M. (2012). Data Mining for the Masses. Computer Global Text Project.https://doi.org/10.1016/j.jspd.2013.01.003

Novia. (2017). Pengertian Telemarketing dan 5 Strategi Suksesnya. Diakses dari http://www.jurnal.id/blog/2017-pengertian-telemarketing-dan-5-strategi-suksesnya/ tanggal 26/10/2019.

Oktanisa, Irvi dan Ahmad Afif Supianto. (2018). Perbandingan Teknik Klasifikasi Dalam Data Mining Untuk Bank Direct Marketing. *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIIK)*, Vol. 5, No. 5, Oktober, hlm. 567-576.

Pal, M. (2005). Random forest classifier for remote sensing classification, *Internatonal Journal of Remote Sensing*, 26:1, 217-222.

Quinlan, J. R. (1987). Simplifying decision trees, *International Journal of Man-Machine Studies* 27, pp: 221-234.

S. Moro, P. Cortez and P. Rita. A Data-Driven Approach to Predict the Success of Bank Telemarketing. Decision Support Systems, Elsevier, 62:22-31, June 2014.

Saez, J. A., Derrac, J., Luengo, J., & Herrera, F. (2014). Statistical computation of feature weighting schemes through data estimation for nearest neighbor classifiers, *Pattern Recognition*, 47(12), 3941-3948 https://doi.org/10.1016/j.patcog.2014.06.012

Saez, J. A., Galar, M., Luengo, J., & Herrera, F. (2013). Tackling the problem of classification with noisy data using Multiple Classifier Systems: Analysis of the performance and robutsness. *Information Sciences*, 247 (June), I-20.https://doi.org/10.1016/j.ins.2013.06.002

Sebastiani, F. (2002). Machine learning in automated text categorization. *ACM Computing Surveys*, 34(1), 1-47.https://doi.org/10.1145/505282.505283

Shearer, Colin. (2000). The CRISP-DM Model: The New Blueprint for Data Mining. *Journal of Data Warehousing, Vol. 5, No. 4 Fall 2000.*

Shmilovici, A. (2009). Support vector machine, In:Maimon O., Rokach L. (eds) Data mining and knowledge discovery handbook, Springer, Boston, MA.

Suntoro, Joko. (2019). Data Mining Algoritma dan Implementasi dengan Pemrograman PHP. Jakarta: PT. Elex Media Komputindo.

Suthaharan, S. (2016). Support vector machine, In: Machine leaning models and algorithms forbig data classification, Integrated Series In Information Systems, vol.36, Springer, Boston, MA.

Suyanto. (2011). Artificial Intelligence - Searching-Reasoning-Planning

- Learning (Edisi Revisi, cetakan ke-2). Bandung : Informatika.

Trevino, Victor and Francesco Falciani. (2006). "GALGO: an R Package for Multivariate Variable Selection Using Genetic Algorithms," *Bioinformatics*, vol. 22, p. 1154–1156.

Zaidi, Nayyar, d.k.k. (2013). "Alleviating Naive Bayes Attribute Independence Assumption by Attribute Weighting," *Journal of Machine Learning Research*, vol. 1, no. 14, pp. 1947 – 1988.

Zukhri, Zainudin. (2014). Algoritma Genetika Metode Komputasi Evolusioner untuk Menyelesaikan Masalah Optimasi. Yogyakarta: Andi Offset.

http://dx.doi.org/10.30649/japk.v10i1.71