

Feature Selection dengan Komparasi Algoritma untuk Prediksi Telemarketing Bank

Fari Katul Fikriah¹, Nur Hayati²

^{1,2}Fakultas Sains dan Teknologi, ITS NU Pekalongan
Jl. Karangdowo No. 9 Kedungwuni Kab. Pekalongan, Jawa Tengah, Indonesia

¹ farichatulfikriyah45@gmail.com

² nurhayatitris@gmail.com

Abstract

Persaingan dalam dunia perbankan membuat pemicu terbentuknya suatu strategi yang dapat menarik nasabah sebanyak-banyaknya. Selain pemasaran dilakukan secara tatap muka kepada calon nasabah, bank dapat menggunakan atau memanfaatkan teknologi untuk melakukan pemasaran dalam jarak jauh seperti menggunakan fax, telepon maupun alat media lain seperti pemasaran lewat telepon (Telemarketing). Oleh karena itu, penting bagi perusahaan untuk memprediksi seberapa tinggi probabilitas para pelanggan agar dapat menerima penjualan atau penawaran berdasarkan jenis karakteristik maupun perilaku calon nasabah selama bertransaksi. Penelitian ini bertujuan untuk meningkatkan akurasi profil nasabah pada suatu perusahaan perbankan melalui klasifikasi serta mengidentifikasi kelompok nasabah yang memiliki probabilitas tinggi untuk berlangganan deposito berjangka panjang dengan menggunakan algoritma Naive Bayes, Logistic Function dan Decision Tree. Sejumlah 45.211 data dari Bank Marketing Data Set diperoleh dari UCI Machine Learning Repository memiliki atribut sebanyak 17 dan bersifat multivariate diimplementasikan pada metode Feature Selection dengan algoritma Decision Tree dimana Feature Selection disini memiliki fungsi mengurangi dimensi, menghilangkan yang tidak relevan dan fitur yang berlebihan, memfasilitasi pemahaman data, mengurangi jumlah data serta dapat meningkatkan akurasi pada algoritma. Hasil menunjukkan bahwa algoritma tersebut mampu dengan optimal untuk memprediksi keputusan klien telemarketing bank dengan hasil akurasi sebesar 91,14% dan dapat diterapkan pada atribut klien bank.

Keywords: Feature Selection, Decision Tree, Telemarketing Bank, Pelanggan

1. Pendahuluan

Bank merupakan lembaga keuangan yang meminjamkan, menerima serta menyimpan uang, membangun kesenjangan antara pemberi pinjaman dan peminjam [1] [2]. Para ekonom mendefinisikan bank sebagai perantara keuangan dimana sebuah lembaga yang bertugas mengumpulkan uang dari menabung, mentransfer hingga meminjamkan. Bank berbeda dengan perantara keuangan lain seperti reksadana atau perusahaan asuransi, hanya dalam jenis kewajiban tertentu yang dapat diterbitkan dan diterima. Perbedaan tersebut sering dianggap sebagai analisis sekunder dan paling standar dari kelompok sistem keuangan bank dan lembaga keuangan lainnya [3]. Telemarketing merupakan sebuah teknik pemasaran guna menghasilkan prospek bisnis yang berkualitas untuk organisasi atau usaha dan teknik cepat guna terhubung dengan klien baru secara pribadi, untuk mengatur janji dan membuat kesepakatan melalui telepon. Pemasaran langsung mempromosikan produk yang dimiliki dengan penjualan tinggi di pasar, teknik seperti ini sering digunakan dalam di dunia perbankan, industri asuransi dan juga *retail*. Telemarketing digunakan guna bisa mendapatkan potensi terbaik dan mendapatkan klien atau prospek untuk sebuah produk atau layanan yang dimiliki sekaligus dapat membantu perusahaan atau sebuah usaha maupun organisasi tertentu untuk dapat lebih banyak mendapatkan keuntungan dengan pemasaran yang dilakukan dengan jarak jauh. Teknik pemasaran dalam bentuk telemarketing ini dapat bermanfaat meyakinkan klien dari jarak jauh dan bersepakat melalui telpon [4].

Telemarketing bank merupakan promosi yang dilakukan perbankan guna menarik nasabah melalui telepon. Persaingan dalam dunia perbankan mengharuskan bank memiliki strategi yang cukup baik dalam menarik nasabah sebanyak-banyaknya. Selain pemasaran dilakukan secara tatap muka kepada calon nasabah, bank dapat menggunakan atau memanfaatkan teknologi untuk melakukan pemasaran dalam jarak jauh seperti menggunakan fax, telepon maupun alat media lain. Telemarketing bank tersebut biasanya dipergunakan bank untuk menawarkan produk dan jasa yang dimiliki oleh bank [5]. Database dari calon pelanggan didapat melalui pemasaran yang telah dilakukan, saat ini banyak bank telah mengadopsi pendekatan prediktif atau analisis bisnis yang menggunakan teknik data mining untuk mengklasifikasi pelanggan potensial sebelum mereka berinventasi. Banyak model telah diusulkan bahwa memiliki kelebihan dan kekurangan pada teknik data mining [6]. Data mining merupakan proses logis yang digunakan untuk mencari data yang berguna melalui jumlah data yang relative besar. Tujuan dari teknik ini adalah guna menemukan pola yang belum diketahui [7]. Data mining sendiri merupakan ekstraksi eksplisit dari informasi yang belum diketahui serta berpotensi untuk membangun program komputer yang menyangring database secara otomatis dan mencari keteraturan atau pola [8].

Data mining memiliki beberapa langkah penyelesaian sehingga menjadi sebuah kesimpulan untuk pengambilan keputusan, salah satu langkahnya ada *preprocessing data*. *preprocessing data* merupakan langkah mengolah data mentah menjadi data yang lebih berguna, efisien dan meningkatkan performa yang lebih maksimal. Penelitian ini menggunakan *preprocessing data* yang dilakukan dengan menggunakan *Feature Selection* yang memiliki fungsi untuk memilih subset dari fitur dengan cara menghilangkan fitur dengan sedikit prediksi informasi. Manfaat dari seleksi fitur adalah mengurangi dimensi, menghilangkan yang tidak relevan, memfasilitasi pemahaman data, mengurangi jumlah data serta dapat meningkatkan akurasi pada algoritma [9] [10]. Seleksi fitur juga merupakan bagaimana mendapatkan subset optimal dari fitur yang relevan dan tidak berlebihan yang akan memberikan solusi optimal tanpa meningkatkan kompleksitas permodelan atau algoritma [11]. Algoritma yang digunakan dalam penelitian juga harus memberikan performa terbaik sehingga menghasilkan akurasi (confusion matrix) yang maksimal pula.

Pada penelitian sebelumnya yang sudah dilakukan oleh Sergio Moro, *et.al* [12] dalam jurnal tersebut digunakan metode *Decision Support System* yang didasarkan dengan model *data-driven* dengan dataset yang sama. Pada penelitian tersebut menggunakan tahap permodelan dengan mengurangi 22 fitur. Dan membandingkan empat model data mining yaitu, *Decision Tree*, *Logistic Regresi*, *Neural Network* dan *SVM* atau *Support Vector Machine*. Penelitian ini juga menyertakan dua metric yakni AUC dan LIFT. Tingkat keberhasilan atau akurasi yang didapatkan setelah melakukan hasil evaluasi *Neural Network* terbaik dengan hasil AUC 0,8 dan LIFT 0,7 adalah sebesar 79%.

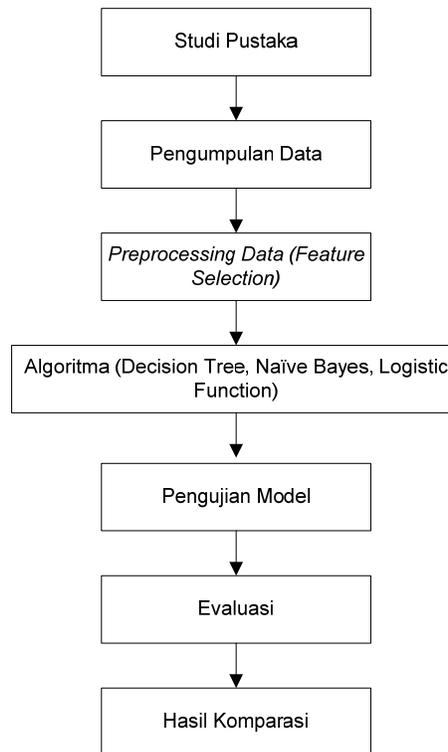
Pada penelitian lain [13] database calon pelanggan yang berasal dari database pemasaran langsung. Pada penelitian tersebut mendeskripsikan bahwa penting bagi perusahaan untuk memprediksi sekumpulan pelanggan dengan probabilitas tinggi untuk menerima penjualan atau penawaran berdasarkan jenis karakteristik maupun perilaku calon nasabah selama melakukan perbelanjaan. Pada penelitian ini berfokus pada membantu bank untuk meningkatkan akurasi profil nasabah melalui klasifikasi serta mengidentifikasi kelompok nasabah yang memiliki probabilitas tinggi untuk berlangganan deposito berjangka panjang dengan menggunakan algoritma *Naïve Bayes*, *Random Forest* dan *Decision Tree*. Penelitian dilakukan untuk mengukur persentase akurasi, presisi serta *recall*. Akurasi tertinggi diperoleh dari algoritma *Decision Tree* dengan jumlah akurasi sebanyak 90%. Venkatesh Yadav, *et. al* [4] menyebutkan pertumbuhan penggunaan teknologi baru serta persaingan yang ketat membuat bank harus mengadopsi pendekatan terbaru untuk dapat menghasilkan kepuasan pelanggan. penelitian ini menyebutkan atau menggunakan berbagai algoritma seperti *Random Forest*, *Rep Tree*, *Naïve Bayes*, *Decision Tree* untuk meningkatkan jumlah pelanggan yang berlangganan. Dari beberapa algoritma *Decision Tree* mendapatkan akurasi yang lumayan tinggi yaitu sebesar 93%.

Penelitian ini memiliki tujuan untuk memprediksi keputusan yang akan dilakukan oleh pengguna atau klien Telemarketing agar memudahkan dalam pengampilan keputusan yang dilakukan oleh Bank serta tujuan lain dari penelitian ini adalah memberikan performa terbaik dalam akurasi, *recall* maupun precision pada prediksi telemarketing bank dengan dataset yang dimiliki tidak mempunyai *Missing Value* atau tidak ada data yang hilang serta karakteristik data

yang dimiliki bersifat *Real* dengan 17 atribut yang dipakai. Pada penelitian ini menggunakan *Feature Selection* dengan algoritma *Decision Tree* yang sebelumnya sudah dikomparasi dengan algoritma lain seperti *Naïve Bayes* dan *Logistic Function*.

2. Metode Penelitian

Penelitian ini menggunakan metode *Feature Selection* dengan komparasi algoritma *Decision Tree*, *Naïve Bayes* dan *Logistic Function*. Pada penelitian sebelumnya dengan menggunakan dataset yang sama dengan menggunakan empat algoritma yakni *Decision Tree*, *Logistic Regresi*, *Neural Network* serta *Support Vector Machine* mendapatkan hasil akurasi 79%. Maka penelitian ini menggunakan *Feature Selection* guna meningkatkan akurasi dan performa yang ada pada penelitian. Adapun tahapan yang dilakukan pada penelitian ini adalah sebagai berikut:



Gambar 1. Tahapan Penelitian

Tahapan pada penelitian seperti tertera pada Gambar 1. di atas adalah dengan melakukan studi pustaka terlebih dahulu, yang kemudian peneliti melakukan pengumpulan data, preprocessing data, melakukan pengujian dengan menggunakan model dan evaluasi hingga di dapatkan hasil dari komparasi yang telah dilakukan.

2.1 Pengumpulan Data

Tahapan pada pengumpulan data ini penulis memakai data public dari UCI *Machine Learning Repository* yaitu “*Bank Marketing Data Set*” dengan link <https://archive.ics.uci.edu/ml/datasets/Bank+Marketing> dimana berisi keterangan sebagai berikut:

Tabel 1. Penjelasan Dataset

Karakteristik Dataset	Multivariate
Karakteristik Atribut	Real

Associated Task	Classification
Jumlah Data	45211
Banyak Atribut	17
Missing Value	Tidak ada

Pada tabel 1. tersebut dijelaskan bahwa karakteristik dari dataset ini adalah *multivariate* dengan jumlah data sebanyak 45211 dengan atribut sebanyak 17 dan tidak ada *missing value* pada data. Data marketing bank pada UCI *Machine Learning Repository* tersebut merupakan marketing menggunakan telepon atau biasa dikenal dengan telemarketing dari sebuah lembaga perbankan Portugis dengan atribut sebagai berikut:

Tabel 2. Data Atribut

No.	Atribut	No.	Atribut
1	Age	11	Duration
2	Job	12	Campaign
3	Marital	13	Pdays
4	Education	14	Previous
5	Default	15	Poutcome
6	Housing	16	Emp. Var.rate
7	Loan	17	Cons.price.idx
8	Contact	18	Cons.conf.idx
9	Month	19	Euribor3m
10	Day_of week	20	Nr.employed
		21	Y

Pada tabel di atas telah dijelaskan beberapa atribut yang ada pada penelitian ini, *class* atau target dari data ini ada “y” dengan atribut memiliki arti biner yaitu “apakah klien sudah berlangganan deposito berjangka?”.

2.2 *Tools* yang digunakan

Aplikasi pendukung yang digunakan pada penelitian ini adalah menggunakan Weka guna mengolah data publik yang diambil dari UCI *Machine Learning Repository*.

3. Hasil dan Pembahasan

Berdasarkan uraian serta metodologi yang telah dipaparkan di atas, hasil serta pembahasan dalam penelitian ini menggunakan data dari UCI *Machine Learning Repository* dengan menggunakan Feature Selection dan algoritma *Naive Bayes*, *Decision Tree*, serta *Logistic Function*. Pada penelitian ini menggunakan model *10 fold cross validation* dan *percentage Split* sebesar 66 % untuk mencari hasil berupa *precision*, *recall* beserta *accuracy* dari data yang telah diolah. Dari pengolahan data tersebut juga mendapatkan hasil berupa *Kappa Statistic*, *Mean Absolute Error* dan *Root Mean Squared Error*.

3.1 Hasil Pengujian Menggunakan *Naive Bayes*

Pengujian menggunakan *Naïve Bayes* data telemarketing bank dengan metode sesuai dengan yang tertera pada pembahasan sebelumnya mendapatkan nilai confusion matrix berupa hasil akurasi sebesar 87,28%, sedangkan untuk nilai *Cappa Statistic* sebesar 0,451, nilai untuk *Mean Absolute Error* sebesar 0,14 dan *Root Mean Squared Error* sebesar 0,332. Dari pengujian ini juga memberikan detail akurasi pada setiap *Class* sebagai berikut:

Tabel 3. Detail Akurasi pada Class Menggunakan metode *Naïve Bayes*

Class	TP-Rate	FP-Rate	Precesion	Recall	F-Measure
0	0,905	0,383	0,949	0,905	0,927
1	0,617	0,095	0,453	0,617	0,522

3.2 Hasil Pengujian Menggunakan *Decision Tree*

Pada pengujian menggunakan metode *Decision Tree* yang sebelumnya sudah melakukan langkah *preprocessing* menggunakan *Feature Selection* didapatkan nilai confusion matrix berupa hasil akurasi sebesar 91, 14% dengan nilai *Kappa Statistic* sebesar 0,529, besar nilai *Mean Absolute Error* sebesar 0,1135 serta besar *Root Mean Squared Error* sebesar 0, 2593 sedangkan untuk detail akurasi pada tiap *class* sesuai dengan tabel yang tertera di bawah ini:

Tabel 4. Detail Akurasi pada Class menggunakan metode *Decision Tree*

Class	TP-Rate	FP-Rate	Precesion	Recall	F-Measure
0	0,959	0,461	0,942	0,959	0,951
1	0,539	0,041	0,624	0,539	0,578

Pada tabel 2. di atas dijelaskan bahwa *Precision* menggunakan metode *Decision Tree* pada *Class 0* sebesar 0,959 atau 95,9% sedangkan untuk *Class 1* sebesar 0,624 atau 62,4%.

3.3 Hasil pengujian Menggunakan *Logistic Function*

Selain menggunakan metode *Naïve Bayes* dan *Decision Tree*, penelitian ini juga menggunakan metode untuk menguji dataset yaitu *Logistic Function*, pada pengujian ini mendapatkan nilai confusion matrix berupa hasil akurasi sebesar 91,07% dengan nilai *Kappa Statistic* sebesar 0,469 nilai *Mean Absolute Error* sebesar 0,1223 dan nilai *Root Mean Squared Error* sebesar 0,2508. Hasil detail dari masing-masing *class* dapat dilihat pada tabel di bawah ini:

Tabel 5. Detail Akurasi pada class menggunakan metode *Logistic Function*

Class	TP-Rate	FP-Rate	Precesion	Recall	F-Measure
0	0,973	0,577	0,930	0,973	0,951
1	0,423	0,027	0,663	0,423	0,516

Hasil pengujian menggunakan ketiga metode yang sudah di jelaskan dapat dilihat bahwa hasil terbaik adalah menggunakan metode *Decision Tree* dengan hasil akurasi sebesar 91,14%, untuk

penjelasan perbedaan akurasi pada masing-masing metode dapat dilihat pada tabel di bawah ini:

Tabel 6. Hasil Akurasi Keseluruhan Metode

	<i>Naïve Bayes</i>	<i>Decision Tree</i>	<i>Logistic Function</i>
Akurasi	87,28%	91,14%	91,07%

Berdasarkan pengujian yang telah dilakukan dapat diketahui bahwa algoritma *Decision Tree* berbasis *Feature Selection* dapat diterapkan berdasarkan atribut pada klien bank. Dengan algoritma tersebut mampu dengan optimal untuk memprediksi keputusan klien telemarketing bank.

4. Kesimpulan

Penggunaan algoritma berbasis *Feature Selection* dari analisis yang diperoleh menggunakan WEKA menunjukkan peningkatan akurasi dan performa dari masing-masing metode. Peningkatan akurasi dan performa dapat diketahui dengan membandingkan metode *Decision Tree* yang memiliki nilai akurasi paling tinggi dari pada akurasi yang diperoleh dari metode *Naïve Bayes* dan metode *Logistic Function* yaitu sebanyak 91,14%. Akurasi tersebut didapat dari nilai confusion matrix. Hal tersebut menunjukkan bahwa pihak marketing bank dapat menggunakan metode *Decision Tree* berbasis *Feature Selection* untuk prediksi telemarketing dalam upaya meningkatkan jumlah pelanggan yang berlangganan.

Daftar Pustaka

- [1] D. D. G. Prabhavathi K, "Banking: Definition and Evolution," *International Journal of Scientific & Engineering Research*, vol. 9, n^o 8, pp. 745-753, 2018.
- [2] V. G. Mala M., "Role of Banking Sector in Financial Inclusion," *International Journal of Computational Research and Development*, vol. 1, n^o 1, pp. 172-176, 2016.
- [3] L. Angeles, "On The Nature of Bank," *University Glasgow*, pp. 316-317, 2019.
- [4] M. S. T. R. V. Yadav, "Classification of Telemarketing Data Using Different Classifier Algorithm," *International Journal of Innovative Technology and Exploring Engineering (IJITEE)*, vol. 9, n^o 12, pp. 1300-1307, 2019.
- [5] R. Sulaeheni, "Predikasi Keputusan Klien Telemarketing untuk Deposito pada Bank Menggunakan Algoritma Naive Bayes Berbasis Backward Elimination," *Jurnal Ilmiah ILKOM*, vol. 8, pp. 182-189, 2016.
- [6] A. N. Asuncion, "D. CA: University of California, School of Information and Computer Science," *UCI Machine Learning Repository*, 2012.
- [7] M. R. M. Bharati, "Data Mining Techniques and Application," *Indian Journal of Computer Science and Engineering*, vol. 1, p. 4.
- [8] T. J. K. Ahmed, "Comparative Analysis of Data Mining Classification Algorithms in Type-2 Diabetes Prediction Data Using WEKA Approach," *International Journal of Science and Engineering (IJSE)*, vol. 7, n^o 2, pp. 155-160, 2014.
- [9] T. N. D. Oreski, "Comparison of Feature Selection Techniques in Knowledge Discovery Process," *TEM Journal*, vol. 3, p. 4, 2014.
- [10] S. J. S. Chormunge, "Efficient Feature Subset Selection Algorithm for High Dimensional Data," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 6, p. 4, 2016.
- [11] M. W. Mwadulo, "Review on Feature Selection Method for Classification Tasks," *International Journal of Computer Application Technology and Research*, vol. 5, pp. 395-402, 2016.

- [12] P. C. P. R. S. Moro, "A Data-driven Approach to Predict the Success of Bank Telemarketing," *Decision Support System*, 2014.
- [13] A. M. C. F. S. Palaniappan, "Customer Profiling using Classification Approach for Bank Telemarketing," *International Journal on Informatics Visualization*, vol. 1, pp. 4-2, 2017.
- [14] K. K. J. D. G. Saranya, "Analysis Of Bank Customers and Prediction of Bank Marketing Strategies," *International Journal of advance Science and Technology*, vol. 29, p. 9, 2020.
- [15] B. A. S. Z. A. Abu-Srhan, "Visualization Analysis in Bank Direct Marketing Prediction," *International Journal of Advanced Computer Science and Application*, vol. 10, p. 7, 2019.