Algoritma *k-Nearest Neighbor* Berbasis *Backward Elimination* Pada *Client Telemarketing*

ST. Aminah Dinayati Ghani, Nur Salman, Mustikasari
STMIK Dipanegara Makassar
Jalan Perintis Kemerdekaan KM. 9 Makassar, Telp. (0411)587194 – Fax. (0411) 588284
e-mail: dinayati.amy@dipanegara.ac.id, nursalman.halim@dipanegara.ac.id, mustikasalman@yahoo.com,

Abstrak

Telemarketing yaitu memasarkan atau mensosialisasikan produk atau jasa melalui telepon. Menurut banyak ahli pemasaran, penawaran melalui *telemarketing* cenderung mudah diterima, karena sifatnya memang secara personal langsung ke konsumen. Deposito berjangka adalah produk simpanan di bank yang penyetorannya maupun penarikannya hanya bisa dilakukan pada waktu tertentu saja. Dengan adanya tenaga *telemarketing* yang menawarkan produk deposito berjangka maka diharapkan dapat meningkatkan efektifitas promosi sehingga meningkatkan jumlah nasabah yang membuka deposito berjangka. Penelitan ini bertujuan untuk memprediksi peningkatan nasabah deposito berjangka melalui jasa *telemarketing* dengan menggunakan Algoritma *k-Nearest Neighbor* berbasis *Backward Elimination*. Hasil dari penelitian ini menunjukkan akurasi algoritma *k-Nearest Neighbor* pada k = 11 adalah 88,25 %, sedangkan akurasi algoritma *k-Nearest Neigbour* berbasis *Backward Elimination* pada k = 9 adalah 89,43 %. Penggunaan algoritma *k-Nearest Neighbor* berbasis *Backward Elimination* lebih meningkat 1,18%, sehingga penawaran melalui *telemarketing* dapat diterapkan untuk meningkatkan nasabah deposito berjangka.

Kata Kunci *- Telemarketing,* Deposito, Fitur Seleksi, Klasifikasi, *Backward Elimination, k-Nearest Neighbor*

Abstract

Telemarketing is marketing or socializing products or services by telephone. According to many marketing experts, offers through telemarketing tend to be easily accepted, because they are personally direct to consumers. Time deposits are savings products at banks whose deposits and withdrawals can only be made at certain times. With the presence of telemarketing staff offering time deposit products, it is expected to increase the effectiveness of promotions, thereby increasing the number of customers who open time deposits. This research aims to predict the increase in time deposits customers through telemarketing services using the k-Nearest Neighbor algorithm based on Backward Elimination. The results of this study indicate the accuracy of the k-Nearest Neighbor algorithm at k = 11 is 88.25%, while the accuracy of the k-Nearest Neigbour based Backward Elimination algorithm at k = 9 is 89.43%. The use of the Backward Elimination k-Nearest Neighbor algorithm is increased by 1.18%, so that offers through telemarketing can be applied to increase customers' time deposits.

Keywords *- Telemarketing, Deposits, Selection Features, Classification, Backward Elimination, k-Nearest Neighbor*

1. Pendahuluan

Globalisasi yang selalu ditandai dengan perkembangan teknologi, terutama teknologi komunikasi, membawa konsekuensi pada manusia untuk selalu meningkatkan kualitas hidupnya. Seiring dengan perkembangan teknologi yang semakin cepat, kebutuhan Sumber Daya Manusia (SDM) dibidang industri komunikasi semakin dibutuhkan. *Telemarketing* berkembang seiring dengan perkembangan teknologi. *Telemarketing* yaitu memasarkan atau mensosialisasikan produk atau jasa melalui telepon.

Menurut banyak ahli pemasaran, penawaran melalui *telemarketing* cenderung mudah diterima, karena sifatnya memang secara personal langsung ke konsumen. Dengan kemajuan teknologi komunikasi, penggunaan telepon tidak hanya sebagai alat komunikasi saja tetapi juga sebagai alat pemasaran. Selain menghemat biaya, masih banyak keuntungan lain yang bisa diperoleh dari penggunaan telepon sebagai alat pemasaran dalam bisnis. Perusahaan mempelajari bahwa biaya yang dikeluarkan seorang *telemarketing* untuk menjual barang keluar kantor jauh lebih efektif dibandingkan dengan tenaga pemasaran yang menjual langsung dilapangan.

Telemarketing lebih cenderung menyampaikan informasi produk dan layanan yang dimiliki perusahaan kepada pelanggan yang kemungkinan akan menggunakan produk atau layanannya. *Telemarketing* menciptakan hubungan antara perusahaan dan calon pelanggan. Dalam proses pemasaran ini akan terjadi interaksi saling mengenal antara perusahaan dan calon pelanggan. Dari interaksi ini calon pelanggan dapat menjadi prospek perusahaan.

Sebagaimana halnya badan usaha yang berorientasi profit, bank juga berupaya menawarkan berbagai produk atau jasa kepada masyarakat semenarik mungkin, antara lain dalam bentuk aneka ragam deposito. Deposito adalah setiap jumlah uang yang dapat disetor oleh seseorang debitur atau penyewa sebagai uang panjar atau uang muka, baik telah dikredit maupun akan dikredit kepadanya atas nama deposito atau uang muka, baik jumlah tersebut akan telah dibayar kepada kreditur atau pemilik atau seseorang lainnya, atau akan telah dilunaskan melalui pembayaran uang atau transfer atau melalui penyerahan barang-barang atau dengan cara lain[1]. Sedangkan deposito berjangka adalah simpanan yang penarikannya hanya bisa dilakukan pada waktu tertentu sesuai dengan tanggal yang telah diperjanjikan antara deposan dan pihak bank. Mengingat simpanan uang atau dana hanya bisa dicairkan ketika jatuh tempo oleh pihak yang namanya tercantum dalam bilyet deposito sesuai dengan tanggal jatuh temponya, maka deposito ini merupakan simpanan atas nama baik perseorangan maupun lembaga, yang artinya di dalam bilyet deposito tercantum nama perorangan atau nama lembaga pemilik deposito berjangka [2].

Untuk itu, suatu bank harus mengambil kesempatan ini dengan melakukan promosi dan strategi pemasaran yang efisien, salah satunya dengan melakukan pemasaran langsung. Salah satu cara yang dapat digunakan agar pemasaran berjalan efisien yaitu dengan memprediksi nasabah yang berpotensi membuka simpanan deposito pada bank tersebut. Prediksi tersebut dapat digunakan dengan menggunakan data-data nasabah yang sudah ada lalu diproses sehingga menemukan hubungan yang berarti, pola, dan kecenderungan [5dengan memeriksa dalam sekumpulan besar data yang tersimpan dalam penyimpanan (*database*) dengan menggunakan teknik pengenalan pola seperti statistik dan matematika. Proses tersebut dinamakan *Data Mining*. *Data mining* adalah suatu proses yang bertujuan untuk menemukan pola secara otomatis atau semi otomatis dari data yang sudah ada di dalam basis data yang dimanfaatkan untuk menyelesaikan suatu masalah [3]. *Data mining* memiliki beberapa teknik, diantaranya *clustering* dan klasifikasi. *Clustering* mengelompokkan objek atau data berdasarkan kemiripan antar data, sehingga anggota dalam satu kelompok memiliki banyak kemiripan dibandingkan dengan kelompok lain. Teknik klasifikasi adalah teknik pembelajaran yang digunakan untuk memprediksi nilai dari atribut kategori target. Klasifikasi bertujuan untuk membagi objek yang ditugaskan hanya ke salah satu nomor kategori yang disebut kelas [3].

Ada beberapa algoritma klasifikasi *data mining* yang dapat digunakan untuk strategi pemasaran dan promosi, seperti yang ditulis oleh Sergio Moro dan Raul M.S. Laureano diantaranya *Naïve Bayes (NB), Decision Trees* (DT), *k-Nearest Neigbor* (k-NN) dan *Support Vector Machines* (SVM) [4]. *k-Nearest Neighbor* (k-NN) termasuk kelompok *instance-based learning*. Algoritma ini juga merupakan salah satu teknik *lazy learning*. *k-Nearest Neigbor* dilakukan dengan mencari kelompok k objek dalam *data training* yang paling dekat (mirip) dengan objek pada data baru atau *data testing*. Penelitian ini akan membahas mengenai penerapan algoritma *k-Nearest Neighbor* untuk peningkatan nasabah deposito berjangka melalui jasa *telemarketing*.

Pemilihan atribut terbaik dapat dilakukan dengan menggunakan metode seleksi fitur. Tujuan seleksi fitur adalah untuk mengidentifikasi fitur atau atribut dalam kumpulan data yang sama pentingnya, dan membuang semua atribut lain yang tidak relevan. *Backward Elimination* adalah metode seleksi fitur yang dilakukan dengan cara pemilihan ke depan yakni menguji semua atribut kemudian menghapus atribut-atribut yang dianggap tidak relevan. Semua atribut diproses satu per satu, jika atribut dianggap tidak berpengaruh atau tidak relevan dalam model maka akan dihapus dari model [5].

Maxsi Ary dan Dyah Ayu Feby Rismiati [5] melakukan penelitan tentang Ukuran Akurasi Klasifikasi Penyakit *Mesotheliom*a Menggunakan Algoritma *k-Nearest Neighbor* dan *Backward*

Elimination. Pengukuran tingkat akurasi yang dilakukan untuk menentukan tindakan selanjutnya, seperti untuk mengetahui deteksi awal pada penyakit *Mesothelioma*. *Mesothelioma* merupakan kanker langka yang mempengaruhi dinding sel tipis dari organ dan struktur internal tubuh manusia yang dapat ditemukan di pleura, peritoneum, dan jantung. Algoritma k-*Nearest Neighbor* digunakan untuk mengklasifikasi objek. *Backward Elimination* dipakai untuk pemilihan atribut yang paling relevan pada proses klasifikasi. Proses seleksi fitur menggunakan *Backward Elimination* dilakukan bersamaan dengan proses pemodelan menggunakan k-*Nearest Neighbor* untuk menemukan subset fitur (set atribut) yang paling relevan. Simpulan dari penelitian ini adalah tingkat akurasi yang dihasilkan oleh algoritma k-*Nearest Neighbor* dalam klasifikasi penyakit mesothelioma sebesar 93.85% dan nilai AUC sebesar 0.995 (termasuk dalam kategori *Excellent Classification*). Sedangkan untuk hasil klasifikasi algoritma *k-Nearest Neighbor* setelah dilakukan seleksi fitur menggunakan *Backward Elimination* menunjukan peningkatan akurasi sebesar 4,61%, sehingga tingkat akurasi yang dihasilkan sebesar 98.46% dan nilai *AUC* sebesar 0.999 (termasuk dalam kategori *Excellent Classification*)

2. Landasan Teori

2.1. Feature Selection

Fitur Seleksi (*feature selection*) merupakan suatu kegiatan yang umumnya bisa dilakukan secara *preprocessing* dan bertujuan untuk memilih fitur yang berpengaruh dan mengesampingkan fitur yang tidak berpengaruh dalam suatu kegiatan pemodelan atau penganalisaan data. Seleksi fitur dipandang sebagai metode pengolahan awal data yang penting dalam tahap pengenalan pola [6].

Ada dua pendekatan dalam *feature selection*, yaitu pendekatan *filter* dan pendekatan *wrapper*. Dalam pendekatan *filter*, setiap fitur dievaluasi secara *independen* sehubungan dengan label kelas dalam *training set* dan menentukan peringkat dari semua fitur, dimana fitur dengan peringkat teratas yang dipilih. Pendekatan *wrapper* menggunakan metode pencarian kecerdasan buatan klasik seperti *greedy hill-climbing* atau *simulated-annealing* untuk mencari subset terbaik dari fitur, dan secara berulang-ulang mengevaluasi subset fitur yang berbeda dengan *cross validation* dan algoritma induksi tertentu.

Manfaat dari pemilihan fitur ini sendiri adalah :

a. Mengurangi dimensionalitas *feature space*, sehingga mengurangi kebutuhan *storage* dan meningkatkan kecepatan algoritma.
b. Menghapus data *redundant*, fitur yang tidak relevan, atau *noise*
c. Mempercepat waktu *running* algoritma *learning*
d. Mengembangkan dan menambah kualitas data
e. Meningkatkan akurasi model
f. Meningkatkan performa

2.2. Backward Elimination

Metode *Backward Elimination* bekerja dengan mengeluarkan satu per satu variabel prediktor yang tidak signifikan dan dilakukan terus menerus sampai tidak ada variabel prediktor yang tidak signifikan. *Backward Elimination* menghilangkan atribut-atribut yang tidak relevan. Algoritma ini didasarkan pada model regresi linear [7]. Langkah-langkah metode *backward* adalah sebagai berikut :

a. Membuat model dengan meregresikan variabel respon Y dengan semua variabel prediktor.
b. Mengeluarkan satu persatu variabel prediktor dengan melakukan pengujian terhadap parameternya dengan menggunakan *partial Ftest*. Nilai Fparsial terkecil dibandingkan dengan Ftabel :
   1.Jika Fparsial < Ftabel, maka X yang bersangkutan dikeluarkan dari model dan dilanjutkan dengan pembuatan model baru tanpa variabel tersebut.
   2.Jika Fparsial > Ftabel, maka proses dihentikan artinya tidak ada variabel yang perlu dikeluarkan dan persamaan terakhir tersebut yang digunakan/dipilih.

2.3. Klasifikasi

Klasifikasi adalah proses penemuan model atau fungsi yang menggambarkan dan membedakan kelas data atau konsep yang bertujuan agar bisa digunakan untuk memprediksi kelas dari objek yang label kelasnya tidak diketahui [7].

Algoritma klasifikasi yang banyak digunakan secara luas, yaitu *Decision/classification trees*, *Bayesian classifiers/Naïve Bayes*, *Neural Networks*, Analisa Statistik, Algoritma Genetika, *Rough sets*, *k-Nearest Neighbor*, Metode *Rule Based*, *Memory based reasoning*, dan *Support Vector Machines (SVM*).

Klasifikasi data terdiri dari dua langkah proses. Pertama adalah *learning (fase training*), dimana algoritma klasifikasi dibuat untuk menganalisa data *training* lalu direpresentasikan dalam bentuk rule klasifikasi. Proses kedua adalah klasifikasi, dimana data tes digunakan untuk memperkirakan akurasi dari rule klasifikasi [7]. Proses klasifikasi didasarkan pada empat komponen [8] :

a. Kelas
Variabel dependen yang berupa kategorikal yang merepresentasikan 'label' yang terdapat pada objek. Contohnya: resiko penyakit jantung, resiko kredit, *customer loyalty*, jenis gempa dan lain-lain.

b. Predictor
Variabel independen yang direpresentasikan oleh karakteristik (atribut) data. Contohnya: merokok, minum alkohol, tekanan darah, tabungan, aset, gaji.

c. Training dataset
Satu set data yang berisi nilai dari kedua komponen di atas yang digunakan untuk menentukan kelas yang cocok berdasarkan *predictor*.

d. Testing dataset
Berisi data baru yang akan diklasifikasikan oleh model yang telah dibuat dan akurasi klasifikasi dievaluasi.

2.4. k-Nearest Neigbor (k-NN)

Algoritma *k-Nearest Neighbor* (k-NN) merupakan salah satu algoritma *supervised learning* di mana hasil klasifikasi data baru berdasarkan pada kategori mayoritas tetangga terdekat ke-k. Algoritma k-NN melakukan klasifikasi tanpa menggunakan model (hanya berdasarkan memori). Dalam klasifikasi, algoritma ini menggunakan mayoritas suara di antara klasifikasi dari k objek [9].

Ketepatan algoritma k-NN ini sangat dipengaruhi oleh ada atau tidaknya fitur-fitur yang tidak relevan, atau jika bobot fitur tersebut tidak setara dengan relevansinya terhadap klasifikasi. Riset terhadap algoritma ini sebagian besar membahas bagaimana memilih dan memberi bobot terhadap fitur, agar performa klasifikasi menjadi lebih baik. Algoritma k-NN ini memiliki konsistensi yang kuat. Ketika jumlah data mendekati tak hingga, algoritma ini menjamin *error rate* yang tidak lebih dari dua kali *Bayes error rate* (*error rate* minimum untuk distribusi data tertentu) [10].

Ada banyak cara untuk mengukur jarak kedekatan antara data baru dengan data lama (*data training*), diantaranya *euclidean distance* dan *manhattan distance* (*city block distance*), yang paling sering digunakan adalah *euclidean distance*, yaitu: [11]

$$\sqrt{(a_1 - b_1)^2 + (a_2 - b_2)^2 + \cdots + (a_n - b_n)^2}$$

Dimana a = a_1,a_2, … ,a_n dan b = b_1,b_2, … ,b_n mewakili n nilai atribut dari dua *record*. Untuk mengukur jarak dari atribut yang mempunyai nilai besar, seperti atribut pendapatan, maka dilakukan normalisasi. Normalisasi bisa dilakukan dengan *min-max normalization* atau *Z-score standardization*. Jika data *training* terdiri dari atribut campuran antara numerik dan kategori, lebih baik gunakan *min-max normalization*. Untuk menghitung kemiripan kasus, digunakan rumus [11] :

$$Similarity\ (p,q) = \frac{\sum_{i=1}^{n} f(p_i, q_i)\ X\ w_i}{w_i}$$

Keterangan :
p = Kasus baru
q = Kasus yang ada dalam penyimpanan
n = Jumlah atribut dalam tiap kasus
i = Atribut individu antara 1 s/d n
f = Fungsi similarity atribut i antara kasus p dan kasus q
w = Bobot yang diberikan pada atribut ke-i

2.5. Cross Validation

Cross-validation merupakan teknik validasi dengan menggunakan seluruh data yang tersedia sebagai *testing* dan *training*. Metode ini memungkinkan proses *training* dan *testing* berulang sebanyak jumlah K dengan data sebanyak 1/K menjadi *testing* dan sisanya sebagai *training* [12]. *Cross Validation* dapat dikatakan sebagai metode untuk membagi data menjadi data *training* dan data *testing.*

2.6. Confusion Matrix

Confusion matrix merupakan metode yang digunakan untuk menentukan akurasi pengenalan klasifikasi. Akurasi yang biasanya dilakukan dengan memilih sampel dari data referensi dan membandingkan kelas pada data referensi ini dengan kelas yang dihasilkan dari klasifikasi sampel [13]. *Confussion Matrik* dapat dilihat pada Tabel di bawah ini :

Tabel 1. Confusion Matrix

Klasifikasi yang benar	Diklasifikasi sebagai	
	+	-
+	*true positive*	*false negative*
-	*false positive*	*true negative*

Keterangan*:*

TP = true positif, yang diklasifikasikan positif
TN = true negatif, yang diklasifikasikan negatif
FP = false positive, yang diklasifikasikan negatif
FN = false negative, yang diklasifikasikan positif

Untuk menghitung tingkat akurasi pada matriks digunakan rumus :

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + TN}$$

Untuk klasifikasi *data mining*, nilai AUC dapat dibagi menjadi beberapa kelompok [14]

0.90 – 1.00 = *Excellent classification*
0.80 – 0.90 = *Good classification*
0.70 – 0.80 = *Fair classification*
0.60 – 0.70 = *Poor classification*
0.50 – 0.60 = *Failure*

3. Metode Penelitian

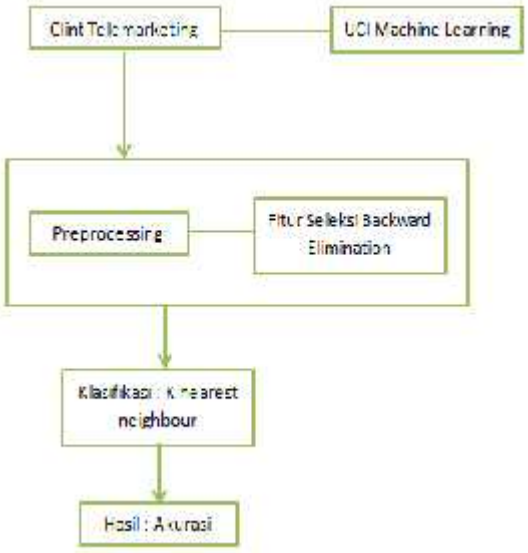
Data yang digunakan pada penelitian ini berasal dari *University of California, Irvine (UCI) Machine Learning* dengan judul Bank Marketing. Dataset ini berjumlah 4521 record dan terdiri dari 17 atribut, dengan 7 atribut bertipe numerik, 6 atribut bertipe kategorikal dan 4 kategori lainnya bertipe binari [15].

Tabel 2. Dataset Bank Marketing

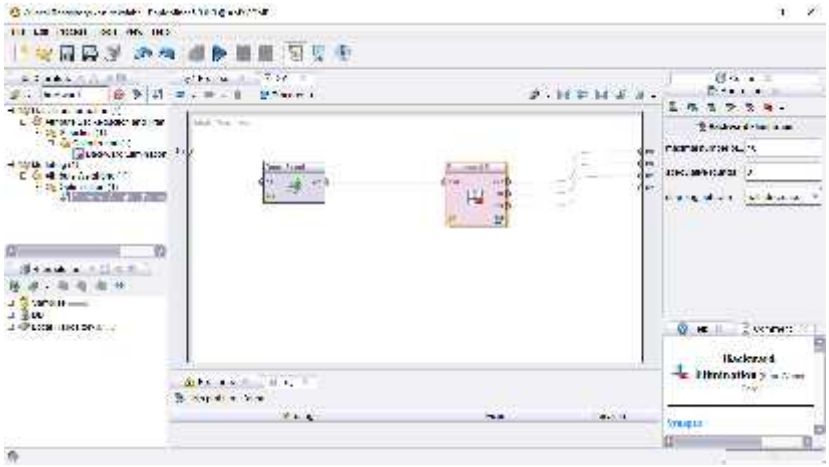
Atribut	Tipe
Age	Numerik
Job	Kategori
Marital	Kategori
Education	Kategori
Default	Binari
Balance	Numerik
Housing	Binari
Loan	Binari

Contact	Kategori
Day	Numerik
Month	Kategori
Duration	Numerik
Campaign	Numerik
Pdays	Numerik
Previous	Numerik
Poutcome	Kategori
Y	Binari

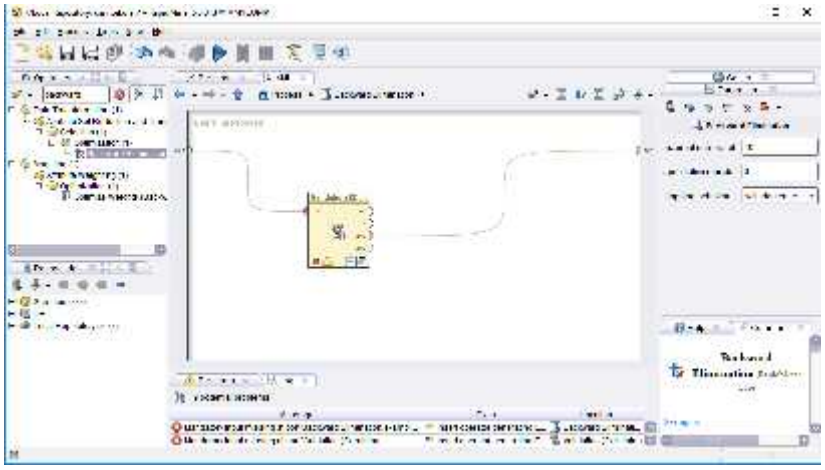
Eksperimen dilakukan dengan melakukan pengujian terhadap dataset serta klasifikasi *client telemarketing* menggunakan *k-Nearest Neighbor* dan menerapkannya dalam *RapidMiner*. Eksperimen akan menghasilkan tingkat akurasi. Adapun langkah-langkah eksperimen ini dapat dilihat pada gambar 1.



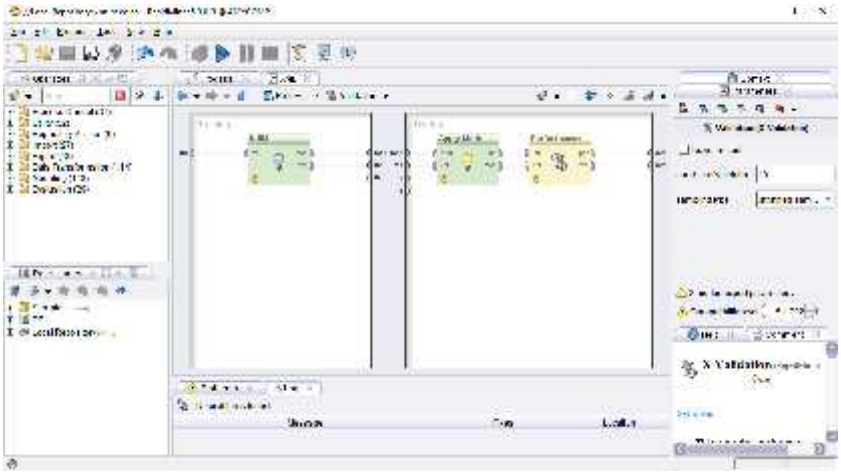
Gambar 1. Desain Eksperimen



Gambar 2. Pembacaan dataset dan seleksi fitur



Gambar 3. Validasi menggunakan *cross validation*



Gambar 4. pencarian nilai k terbaik menggunakan *k-NN*

Pelatihan c*lasifier* membutuhkan waktu, biasanya beberapa kriteria selain dari tingkat pengenalan digunakan untuk pemilihan fitur. Namun hal ini dapat menyebabkan penurunan kemampuan generalisasi dengan pemilihan fitur. Untuk mengatasi masalah ini maka diusulkan pemilihan fitur yang diubah oleh *cross validation.*

Dengan menggunakan *cross validation* akan dapat mengevaluasi kinerja model atau algoritma dimana data dipisahkan menjadi dua subset yaitu data proses pembelajaran dan data validasi atau evaluasi. Model atau algoritma dilatih oleh subset pembelajaran dan divalidasi oleh subset validasi (gambar 3).

4. Hasil dan Pembahasan

Pada percobaan pertama penulis menggunakan algoritma *k-Nearest Neighbor* terlebih dahulu tanpa menggunakan *Backward Elimination* untuk seleksi fitur.

Percobaan tersebut menggunakan 4521 data dan 17 atribut prediktor yang sudah dinormalisasi menjadi data numerik terlebih dahulu. Percobaan tersebut dilakukan beberapa kali karena setiap percobaan mendapatkan nilai akurasi dan tingkat kesalahan yang berbeda.

Tabel 3. Hasil percobaan Algoritma *k-Nearest Neighbor*

k-	Akurasi (%)	Precision	Recall	AUC
1	84,38	31,06	29,72	0,500
3	86,71	37,51	24,17	0,703
5	87,35	39,21	19,17	0,735
7	87,66	40,02	16,87	0,763
9	87,94	42,29	14,95	0,783
11	88,25	45,95	15,13	0,793

Berdasarkan tabel 2 di atas yang merupakan hasil dari percobaan yang dilakukan pada *Rapid Miner*, maka nilai akurasi tertinggi diperoleh pada k-11 dengan nilai akurasi sebesar 88,25 %.

Tabel 4. Hasil percobaan Algoritma *k-Nearest Neighbor* berbasis *Backward Elimination*

k-	Akurasi (%)	Precision	Recall	AUC
1	86,35	39,64	37,61	0,500
3	88,39	49,06	34,56	0,744
5	89,21	54,75	33,40	0,776
7	89,29	56,02	30,69	0,797
9	89,43	57,12	32,45	0,808
11	89,25	56,69	30,13	0,820

Berdasarkan tabel diatas yang merupakan hasil dari percobaan yang dilakukan pada *Rapid Miner*, maka nilai akurasi tertinggi pada k-9 dengan nilai akurasi sebesar 89,43 %.

Setelah melihat perbandingan dari kedua percobaan tersebut maka dapat ditarik kesimpulan bahwa nilai akurasi tertinggi sebesar 89,43 % dicapai pada saat dilakukan percobaan dengan menggunakan algoritma *k-Nearest Neighbor* dengan fitur seleksi *Backward Elimination*, sedangkan nilai akurasi tertinggi sebesar 88,25% dicapai dengan menggunakan algoritma *k-Nearest Neighbor* tanpa menggunakan fitur seleksi *Backward Elimination.*

Penerapan metode fitur seleksi *Backward Elimination* dalam klasifikasi ini berpengaruh terhadap hasil akurasi yang didapatkan, karena dengan menerapkan *Backward Elimination* maka akan menghilangkan atribut-atribut yang tidak relevan. Sehingga untuk klasifikasi penentuan nasabah deposito pada *client telemarketing* menggunakan *k-Nearest Neighbor* ini diperoleh akurasi tertinggi dengan menerapkan fitur seleksi *Backward Elimination* dengan hasil akurasi sebesar 89,43 % dengan nilai *AUC* sebesar 0.80 termasuk dalam kategori *Good Classification.* Penggunaan algoritma *k-Nearest Neighbor* berbasis *Backward Elimination* lebih meningkat 1,18%, sehingga penawaran melalui telemarketing dapat diterapkan untuk meningkatkan nasabah deposito berjangka.

Untuk lebih jelasnya perbandingan dari hasil percobaan di atas dapat di lihat pada gambar 5 berikut ini:

Gambar 5. Grafik perbandingan akurasi hasil percobaan

5. Penutup

Klasifikasi *client telemarketing* dengan menggunakan algoritma *k-Nearst Neighbor* telah dilakukan dengan hasil akurasi 89,43%. Hasil ini diperoleh dengan menggunakan fitur seleksi *Backward Elimination*. Penggunaan algoritma *k-Nearest Neighbor* berbasis *Backward Elimination* lebih meningkat 1,18% dibandingkan jika hanya menggunakan algoritma k-NN saja, *sehingga* penawaran *melalui telemarketing dapat diterapkan untuk meningkatkan nasabah deposito berjangka.* Penggunaan fitur seleksi *Backward Elimination* dalam pemrosesan data akan mempengaruhi hasil pencapaian akurasi yang didapatkan. Beberapa hal masih perlu dilakukan untuk meningkatkan kinerja dari penelitian ini. Salah satunya adalah dengan melakukan pengujian *k-Nearest Neighbor* terhadap data set yang berbeda.

6. Daftar Pustaka

[1] Simorangkir, O. P, Drs , Dasar-dasar dan Mekanisme Perbankan, Aksara Persada Indonesia, Jakarta, 1986.
[2] Alvino Dwi Rachman Prabowo," Prediksi Nasabah yang Berpotensi Membuka Simpanan Deposito Menggunakan Naive Bayes Berbasis Particle Swarm Optimization,"Pascasarjana Universitas Dian Nuswantoro, Semarang.
[3] D. T. Larose, Discovering Knowledge In Data.United States of America: John Wiley & Sons, Inc., 2005.
[4] M. Bramer, Principles of Data Mining, London: Springer, 2013.
[5] Maxsi Ary, Dyiah Ayu Feby Rismiati, "Ukuran Akurasi Klasifikasi Penyakit *Mesotheliom*a Menggunakan Algoritma *K-Nearest Neighbor* dan *Backward Elimination",* SATIN - Sains dan Teknologi Informasi, Vol. 5, No. 1, Juni 2019 ISSN: 2527-9114
[6] S. Moro and R. M. S. Laureano, "Using DataMining for Bank Direct Marketing: An Application of The CRISP-DM Methodology," Instituto Universitário de Lisboa, 2011.
[7] I. Wang, N. A. dan R. I., "Sequential forward selection approach to the non-unique oligonucleotide probe selection problem," Proceedings of the 3rd IAPR International Conference on Pattern Recognition in Bioinformatics, pp. 262-275, 2008.
[8] J. Han dan M. Kamber, "Data Mining: Concepts, Models, and Techniques," dalam Springer, Verlag Berlin Heidelberg, 2011.
[9] F. Gorunescu, Data Mining: Concept, Models and Techniques, Springer, 2011.
[10] Purwanto,"Materi Kuliah Data Mining, Pascasarjana Universitas Dian Nuswantoro, Semarang, 2015.
[11] Leidiyana Henny, "Penerapan Algoritma KNearest Neighbor Untuk Penentuan Resiko Kredit Kepemilikan Kendaraan Bemotor", Thesis MI, STMIK Nusa Mandiri. 2013.

[12] Y. Bengio and Y. Grandvalet, “Bias In Estimating The Variance of K-Fold Cross-Validation,” Learning.
[13] S. V. Stehman, “Selecting And Interpreting Measures Of Thematic Classification Accuracy,” Remote Sens. Environ., vol. 62, no. 1, pp. 77–89, 1997
[14] Sulaehani, R. (2016). Prediksi Keputusan Klien Telemarketing untuk Deposito Pada Bank Menggunakan Algoritma Naive Bayes Berbasis Backward Elimination. Jurnal Ilmiah ILKOM, 182- 189.
[15] Univesity of California, “Machine Learning Repository, ”[online]Avalaible: https//archieve.ics.uci.edu/ Datasets.