

BUSINESS INTELLIGENCE UNTUK PREDIKSI CUSTOMER CHURN TELEKOMUNIKASI

Muhammad Arifin

Program Studi Sistem Informasi, Fakultas Teknik, Universitas Muria Kudus
Gondangmanis, PO Box 53, Bae, Kudus 59352
Email: arifin.m@umk.ac.id

Abstrak

Customer atau pelanggan merupakan hal yang sangat penting bagi kelangsungan hidup suatu perusahaan barang maupun jasa. Dengan ditinggalkan pelanggan sebuah perusahaan akan mengalami kerugian yang sangat besar bahkan untuk memperoleh pelanggan baru memerlukan biaya hingga 10 kali lipat lebih mahal dari biaya untuk mempertahankan pelanggan yang ada. Perusahaan telekomunikasi secara rutin menghasilkan dan menyimpan sejumlah besar data berkualitas tinggi, memiliki basis pelanggan yang sangat besar, dan beroperasi dalam lingkungan yang cepat berubah dan sangat kompetitif serta rentan terhadap customer churn (pelanggan yang meninggalkan perusahaan). Prediksi customer churn muncul sebagai Business Intelligence (BI) yang sangat penting untuk telekomunikasi modern. Penelitian ini melakukan studi literatur untuk mendapatkan informasi tentang permasalahan yang ada selanjutnya eksperimen untuk memprediksi customer churn dalam bidang Telekomunikasi dengan menerapkan algoritma prediksi, menganalisa hasil dan mendokumentasikannya. Dari hasil penelitian yang telah dilakukan dapat diketahui bahwa dengan menerapkan BI pada prediksi customer churn yang menggunakan algoritma KNN menghasilkan tingkat akurasi sebesar 88% pada nilai K 5 keatas. Dengan hasil tersebut diharapkan dapat membantu pihak pengambil keputusan dalam menemukan customer yang memiliki indikasi untuk churn.

Kata kunci: *business intelligence, prediksi, customer churn*

1. PENDAHULUAN

Customer atau pelanggan merupakan hal yang sangat penting bagi kelangsungan hidup suatu perusahaan baik perusahaan barang maupun jasa. Dengan ditinggalkan pelanggan sebuah perusahaan akan mengalami kerugian yang sangat besar selain itu pelanggan merupakan sebagian dari indikasi besar kecilnya sebuah perusahaan, apabila sebuah perusahaan memiliki sedikit pelanggan maka bisa dikatakan perusahaan tersebut kecil sedangkan sebaliknya apabila sebuah perusahaan dikatakan besar dapat dilihat dari jumlah pelanggannya yang jumlahnya besar pula.

Menurut Yu dkk., (2011) *Customer Churn* didefinisikan sebagai kecenderungan pelanggan untuk berhenti melakukan bisnis dengan sebuah perusahaan. Hal ini telah menjadi isu penting yang merupakan salah satu tantangan utama oleh banyak perusahaan di era global ini dan harus dihadapinya. Disamping itu menurut Khakabi dkk.,(2010) untuk memperoleh pelanggan baru itu memerlukan biaya hingga 10 kali lipat lebih mahal dari biaya untuk mempertahankan pelanggan yang ada. Melihat begitu mahalnya untuk memperoleh pelanggan baru tentunya perusahaan akan lebih memilih mempertahankan pelanggan dibanding dengan mendapatkan pelanggan baru. Melihat hasil fakta tersebut maka banyak perusahaan sekarang lebih beralih untuk mempertahankan pelanggan yang ada dan menghindari *churn* pelanggan. Di era modern ini banyak bisnis yang berkembang dari “*product-centered*” ke “*Customer-Centered*” (Coussement dan Poel 2008). Sebelum perusahaan mengetahui ini perusahaan hanya berfokus pada penjualan jasa dan produk dengan sedikit pengetahuan mengenai pelanggan yang membeli produknya.

Integrasi proses pekerjaan di suatu organisasi dengan teknologi informasi pada saat ini sudah menjadi kebutuhan mutlak. Hal ini dikarenakan dengan menggunakan teknologi informasi

proses untuk menganalisa masalah dan mengambil keputusan lebih cepat dan akurat. Kelangsungan hidup suatu organisasi dipengaruhi oleh ketersediaan data dan informasi yang lengkap, benar dan tepat (Depkeu RI., 2007).

Weiss (2009) mengatakan perusahaan telekomunikasi secara rutin menghasilkan dan menyimpan sejumlah besar data berkualitas tinggi, memiliki basis pelanggan yang sangat besar, dan beroperasi dalam lingkungan yang cepat berubah dan sangat kompetitif. Perusahaan telekomunikasi memanfaatkan BI untuk meningkatkan upaya pemasaran mereka, mengidentifikasi penipuan, dan lebih baik mengelola jaringan telekomunikasi mereka. Namun, perusahaan-perusahaan ini juga menghadapi sejumlah tantangan BI karena ukuran besar set data mereka, aspek sekuensial dan temporal data mereka, dan kebutuhan untuk memprediksi kejadian yang sangat langka seperti penipuan pelanggan dan jaringan kegagalan secara real-time. BI dapat dilihat sebagai sarana otomatis menghasilkan beberapa pengetahuan ini langsung dari data.

Dalam perusahaan telekomunikasi, mempertahankan pelanggan lebih mudah dan murah dibandingkan perusahaan harus mencari pelanggan baru, Prediksi *cusromer churn* muncul sebagai *Bussiness Intelligence* (BI) yang sangat penting untuk telekomunikasi modern. Tujuan utama dari prediksi *cusromer churn* adalah untuk memperoleh informasi kemungkinan dari pelanggan yang akan berpindah ke perusahaan lain menggunakan prediksi model. Saat ini menunjukkan interaksi analisa pelanggan dengan cara menilai *customer churn* dari tingkat sosial yang dapat meningkatkan akurasi dari prediksi *customer churn* (Richter dan Slonim, 2010).

Permasalahan diatas mampu dijawab oleh *Business Intelligence* (BI), dimana BI merupakan salah satu bentuk implementasi teknologi informasi yang digunakan untuk membantu kegiatan seperti mengumpulkan data, menyediakan akses, serta menganalisa data dan informasi mengenai kinerja perusahaan. Dengan kegiatan BI tersebut maka sebuah organisasi atau perusahaan akan dengan mudah dalam mengambil keputusan secara cepat dan tepat. Singkatnya BI dapat diartikan sebagai pengetahuan yang didapatkan dari hasil analisis data yang diperoleh dari kegiatan (usaha) suatu organisasi. BI biasanya dikaitkan dengan upaya untuk memaksimalkan kinerja suatu organisasi.

BI dapat membantu suatu organisasi mendapatkan pengetahuan yang jelas mengenai faktor-faktor yang mempengaruhi kinerja organisasi sehingga dapat membantu organisasi dalam pengambilan keputusan serta sekaligus meningkatkan keunggulannya (*competitive advantage*). BI juga dapat membantu suatu organisasi dalam menganalisis perubahan tren yang terjadi sehingga akan membantu organisasi menentukan strategi yang diperlukan dalam mengantisipasi perubahan tren tersebut. Upaya memaksimalkan kinerja organisasi merupakan hal yang prioritas saat ini. Organisasi yang secara jelas mampu mengidentifikasi, menjelaskan, dan mengimplementasikan strateginya akan mampu berkembang dan berkompetisi lebih baik (Depkeu RI., 2007).

2. METODOLOGI

Metode penelitian dalam penelitian ini menggunakan 2 metode yaitu :

1. Metode melalui studi literature.
2. Metode eksperimen dengan menggunakan beberapa tahapan diantaranya:
 - a. Pengumpulan data
 - b. Pemilihan atribut
 - c. Penerapan algoritma
 - d. Evaluasi hasil
 - e. Dokumentasi eksperimen

Adapun skema alur tahapan penelitian ini ditunjukkan pada gambar 1



Gambar 1. Skema Alur Tahapan Penelitian

3. HASIL DAN PEMBAHASAN

3.1. Metode Melalui Studi Literature

Metode melalui studi literature yang bertujuan mendapatkan pengetahuan atau *domain* dari penelitian yang akan dilakukan. Studi literatur tersebut didapatkan melalui berbagai sumber antara lain buku, jurnal, paper, dan sebagainya. Adapun hal-hal yang dapat diambil dari metode ini diantaranya adalah:

- Studi Pendahuluan : tahap ini merupakan kegiatan untuk menemukan informasi tentang obyek permasalahan yang ada. Permasalahan-persalahan yang berkembang beberapa tahun terakhir dalam sebuah organisasi atau perusahaan khususnya mengenai *customer churn*.
- Studi Pustaka : tahapan ini adalah tahap untuk menemukan penelitian-penelitian yang sejenis dengan penelitian ini yang nantinya dijadikan sebagai referensi dan pendukung teori dalam menyelesaikan permasalahan yang diangkat.
- Perumusan Masalah : adapun pada tahapan selanjutnya setelah mendapatkan permasalahan utama dari obyek penelitian yang dilengkapi dasar teori dari studi pustaka yang mendukung maka masalah yang ada dapat dirumuskan dengan baik.

3.2. Metode Eksperimen

Metode eksperimen ini digunakan untuk menganalisa data yaitu memilah label dan variabel yang selanjutnya data digunakan dalam proses prediksi. Proses yang berjalan dalam metode ini menggunakan alat bantu *Rapid Miner 5*, adapun tahapan dalam metode ini adalah sebagai berikut:

3.1.1. Pengumpulan Data

Kegiatan pengumpulan data dapat dilakukan dengan menagambil dari database perusahaan yang digunakan sebagai obyek penerapan BI. Data yang digunakan didalam penelitian ini adalah data telekomunikasi di Colombia dimana dataset *customer churn* diambil dari database-UCI California University. Dalam dataset ini mendefinisikan transaksi panggilan yaitu *churn* per satu pelanggan seluler dari satu perusahaan telekomunikasi, dalam waktu tiga bulan terus menerus. Terdapat 21 fitur. Data *customer churn* ini terdiri dari 5000 *tuple (record)*, terdiri dari 4293 berlabel *false* dan 707 berlabel *true*, terdiri dari 51 negara bagian distrik Colombia.

3.1.2. Pemilihan Atribut

Kegiatan pemilihan atribut digunakan untuk memisahkan antara atribut label (atribut yang akan digunakan sebagai kunci prediksi) dengan atribut variabel prediksi dalam memprediksi sebuah data. Dalam dataset ini atribut yang bernilai *false* atau *true* dipilih sebagai label dan yang lainnya dijadikan sebagai variabel. Pada tabel 1 memperlihatkan atribut-atribut dalam dataset yang digunakan.

Tabel 1. Keterangan Atribut Data Set

Nama Atribut	Keterangan
<i>State</i>	untuk 51 negara bagian <i>District of Columbia</i>
<i>Account Length</i>	berapa lama akun aktif
<i>Area Code</i>	kode area
<i>Phone Number</i>	nomer telepon yang digunakan sebagai ID pelanggan
<i>International Plan</i>	rencana internasional
<i>Voice Mail Plan</i>	rencana pesan suara
<i>Number Vmail Messages</i>	jumlah pesan <i>voice mail</i>
<i>Total Days</i>	total panggilan sehari pada siang hari, yang terdiri dari : <i>total day minutes</i> (jumlah layanan per menit), <i>total day calls</i> (jumlah panggilan) dan <i>total day charge</i> (jumlah biaya)
<i>Total Eve</i>	total panggilan sehari pada sore hari, yang terdiri dari <i>total eve minutes</i> (jumlah layanan per menit), <i>total eve calls</i> (jumlah panggilan) dan <i>total eve charge</i> (jumlah biaya)
<i>Total Night</i>	total panggilan sehari pada malam hari, yang terdiri dari <i>total night minutes</i> (jumlah layanan per menit), <i>total night calls</i> (jumlah panggilan) dan <i>total night charge</i> (jumlah biaya)
<i>Total International</i>	total panggilan yang digunakan untuk panggilan internasional, yang terdiri dari <i>total intl minutes</i> (jumlah layanan per menit), <i>total intl calls</i> (jumlah panggilan) dan <i>total intl charge</i> (jumlah biaya)
<i>Number Customer Service Calls</i>	jumlah panggilan ke layanan pelanggan
<i>Label</i>	indikasi tidak <i>churn</i> dan <i>churn</i> (<i>False</i> and <i>True</i>)

3.1.3. Penerapan Algoritma

Prediksi sebuah data yang dalam hal ini adalah data pelanggan telekomunikasi dapat diprediksi dengan menggunakan beberapa algoritma prediksi diantaranya SVM, Logistik Regresi dan KNN. Dalam penelitian ini algoritma yang digunakan adalah KNN dimana algoritma ini sangatlah sederhana (Harrington, 2012), bekerja berdasarkan jarak terpendek dari *query instance* ke *training sample* untuk menentukan KNNnya. Training sample diproyeksikan keruang berdimensi banyak, dimana masing-masing dimensi merepresentasikan fitur dari data. Ruang ini dibagi menjadi bagian-bagian berdasarkan klasifikasi *training sample*. Sebuah titik pada ruang ini ditandai kelas c jika kelas c merupakan prediksi yang paling banyak ditemui pada k buah tetangga terdekat dari titik tersebut maka sampel tersebut masuk kedalam kelas c. Dekat atau jauhnya tetangga biasanya dihitung berdasarkan persamaan *Euclidean Distance*.

Jarak *Euclidean* paling sering digunakan menghitung jarak (Deepa dan Ladha, 2011). Jarak *euclidean* berfungsi menguji ukuran yang bisa digunakan sebagai interpretasi kedekatan jarak antara dua obyek. yang direpresentasikan pada persamaan 1.

$$D(a,b) = \sqrt{\sum_{k=1}^d (a_k - b_k)^2} \quad \dots(1)$$

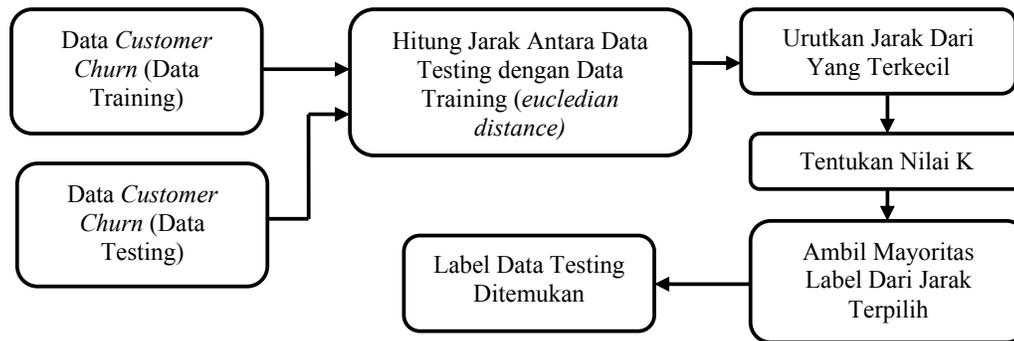
Keterangan:

D(a,b) : jarak skalar dari dua buah vektor a dan b dari matrik berukuran D dimensi

k : data tranning ke n d : jumlah data training

a : data training b : data testing

Persamaan 1 adalah persamaan untuk mencari jarak terdekat antara data testing dengan data taining. Untuk mengetahui proses lengkap dari prediksi *customer churn* telekomunikasi ditunjukkan pada gambar 2.



Gambar 2. Blok Diagram Algoritma Prediksi K-NN

Gambar 2 blok diagram algoritma prediksi *K-NN* diatas memperlihatkan langkah-langkah dalam memprediksi data *customer churn*, dimana antara data training dan data testing dihitung menggunakan persamaan *euclidian distance*, setelah semua data dihitung jaraknya langkah selanjutnya adalah mengurutkan jarak dari jarak terkecil terbesar selanjutnya menentukan nilai *k*, dimana nilai *k* adalah jumlah data terpilih. Dari data terpilih tersebut kemudian dipilih mayoritas labelnya yang sekaligus merupakan hasil dari label data.

Tabel 2. Data Set

State	Account Length	area code	Phone Number	international plan	voice mail plan	number vmail messages	total day minutes	total day calls	total day charge
KS	128	415	382-4657	No	yes	25	265,1	110	45,07
OH	107	415	371-7191	No	yes	26	161,6	123	27,47
NJ	137	415	358-1921	No	No	0	243,4	114	41,38
VT	86	415	373-8058	No	yes	34	129,4	102	22
ID	85	408	350-8884	No	yes	27	196,4	139	33,39

Tabel 3. Lanjutan Tabel 2

total eve minutes	total eve calls	total eve charge	Total night minutes	total night calls	total night charge	total intl minutes	total intl calls	total intl charge	number customer service calls	Churn
197,4	99	16,78	244,7	91	11,01	10	3	2,7	1	False.
195,5	103	16,62	254,4	103	11,45	13,7	3	3,7	1	False.
121,2	110	10,3	162,6	104	7,32	12,2	5	3,29	0	False.
267,1	104	22,7	154,8	100	6,97	9,3	16	2,51	0	False.
280,9	90	23,88	89,3	75	4,02	13,8	4	3,73	1	?

Pada tabel 2 dan tabel 3 data terahir belum diketahui apakah pelanggan tersebut masuk kedalam kategori *true (churn)* atau *false (not churn)*, untuk mengetahui label dari data tersebut

maka harus dihitung jaraknya menggunakan persamaan 1. Data yang sudah dihitung selanjutnya di urutkan dari jarak data yang paling kecil kepada yang paling besar nilai jaraknya, selanjutnya ditentukan nilai K-nya, dari proses itu maka akan diketahui label dari data tersebut dari mayoritas label yang didapat.

3.1.4. Evaluasai Hasil

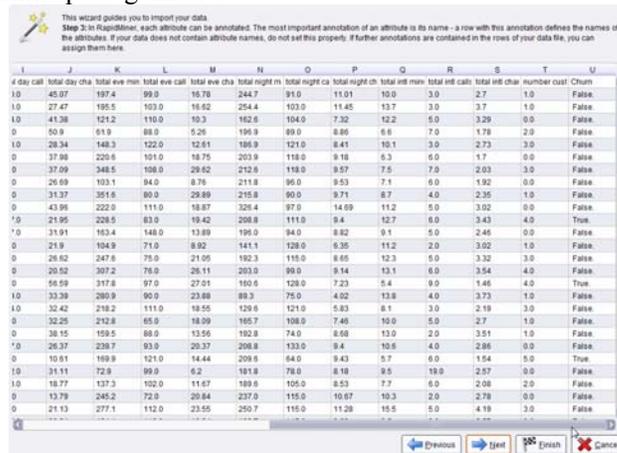
Dari berbagai eksperimen yang dilakukan dalam memprediksi *customer churn* Telekomunikasi dengan menggunakan Algoritma KNN diperoleh data sebagai berikut :

Tabel 4. Hasil Prediksi Customer Churn Telekomunikasi

Prediksi	K															
	1		3		5		7		9		11		13		15	
	Acc	Auc	Acc	Auc	Acc	Auc	Acc	Auc	Acc	Auc	Acc	Auc	Acc	Auc	Acc	Auc
KNN	82	0,5	86,7	0,68	88,3	0,69	88,4	0,69	88,6	0,69	88,3	0,7	88,3	0,7	88,2	0,71

3.1.5. Dokumentasi Eksperimen

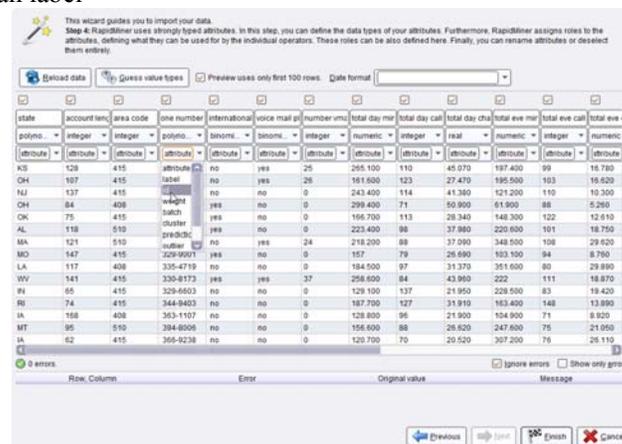
1. Mengambil data dari file excel yang merupakan data *customer churn* dari *uci dataset* adapun prosesnya dapat dilihat pada gambar 3.



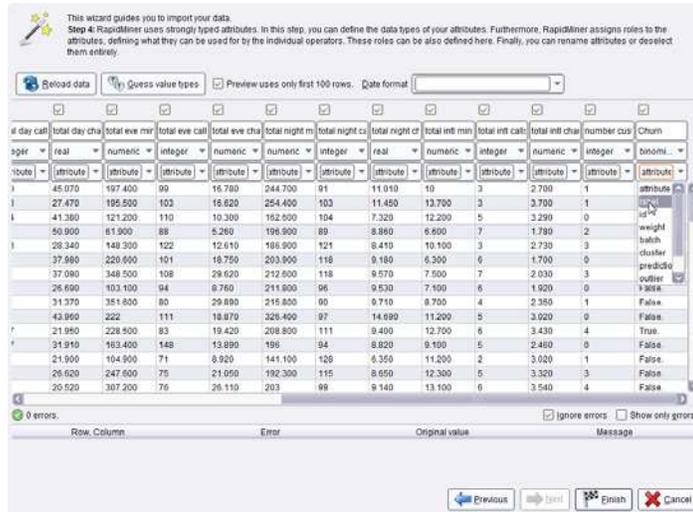
Gambar 3. Import Data Dari Dataset

Pada gambar 3 menunjukkan data *chustomer churn* yang telah di *import* kedalam *tool Rapidminer*.

2. Menentukan id dan label



Gambar 4 Pemilihan ID

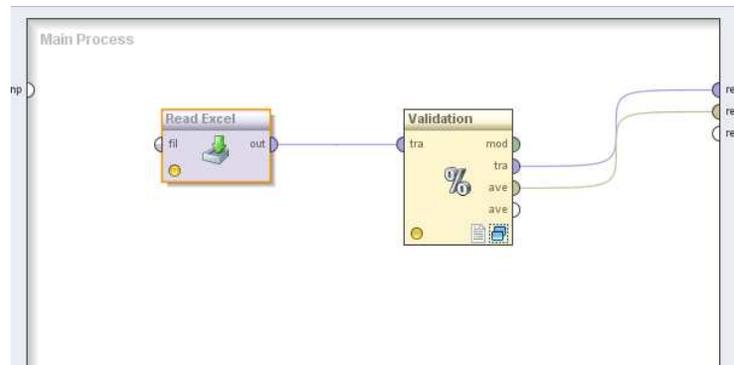


Gambar 5. Penentuan label

Pada gambar 4 dan 5 memperlihatkan proses pengambilan id sebagai kunci utama (*primary key*) pada data *customer churn* telekomunikasi dan pengambilan label sebagai kelas dalam memprediksi *customer churn*.

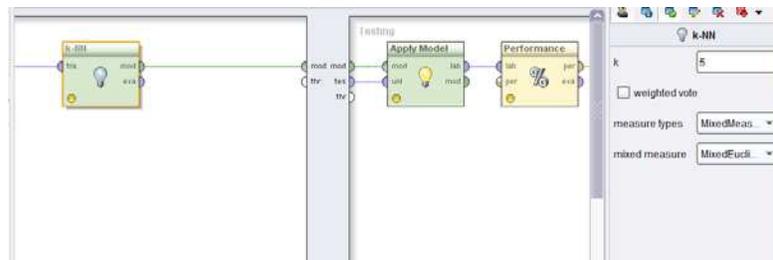
3. Penerapan Algoritma KNN dan Penentuan Nilai K

Proses penerapan algoritma dilakukan setelah data diambil dari database disamping itu data harus sudah ditentukan id dan labelnya untuk diprediksi. Gambar 6 menunjukkan penerapan algoritma KNN dalam memprediksi *customer churn*.



Gambar 6. Penerapan Algoritma

Selanjutnya memilih nilai K yang paling tepat untuk memperoleh nilai akurasi yang paling baik.



Gambar 7 Setting Nilai K

Pada gambar 6 merupakan proses prediksi menggunakan algoritma K -NN yang di validasi menggunakan 10 cros validasi, sedangkan pada gambar 7 memperlihatkan seting nilai k , nilai k ini nantinya akan mempengaruhi akurasi prediksi .

4. KESIMPULAN

Dari hasil penelitian ini dapat disimpulkan bahwa penerapan BI untuk mengetahui prediksi *customer churn* dengan menggunakan algoritma prediksi KNN dapat diterapkan didunia telekomunikasi. Penerapan BI dapat membantu pengambilan keputusan bagi tingkat elit pengambil keputusan dengan cepat dan akurat dengan tingkat akurasi 88% dimana proses pengambilan keputusan didapat dari hasil analisa yang dilakukan. Sebagai contoh pengambil keputusan adalah apabila suatu pelanggan ataupun kelompok pelanggan tertentu yang diketahui sejak dini bahwa pelanggan tersebut cenderung akan meninggalkan perusahaannya maka dapat ditahan dengan berbagai cara diantaranya adalah dengan memberikan promosi-promosi yang menarik dan lain sebagainya.

Penelitian ini mendapatkan hasil 88%, diharapkan penelitian mendatang dapat melanjutkan dengan menambah algoritma pemilihan fitur, imbalant data dan algoritma lain dengan tujuan untuk memperoleh hasil akurasi yang lebih baik.

5. DAFTAR PUSTAKA

- Dep. Keu. RI, "Laporan Tim Studi Tentang Implementasi Business Intelligence" pp. 1-2, Des. 2007.
- K. Coussement and D. Van den Poel, "Integrating the voice of customers through call center emails into a decision support system for churn prediction," *Information & Management*, vol. 45, no. 3, pp. 164–174, Apr. 2008.
- M. Weiss, "Data Mining in the Telecommunications Industry," *IGI Global, Fordham University, USA*, 2009
- P. Harrington, *Machine Learning in Action*. USA: Manning Publications, 2012, p. 18.
- S. KhakAbi, M. R. Gholamian, and M. Namvar, "Data Mining Applications in Customer Churn Management," *2010 International Conference on Intelligent Systems, Modelling and Simulation*, pp. 220–225, Jan. 2010.
- T. Deepa and L. Ladha, "Feature Selection Methods And Algorithms," *International Journal on Computer Science and Engineering (IJCSSE)*, vol. 3, no. 5, pp. 1787–1797, 2011.
- X. Yu, S. Guo, J. Guo, and X. Huang, "An extended support vector machine forecasting framework for customer churn in e-commerce," *Expert Systems with Applications*, vol. 38, no. 3, pp. 1425–1430, Mar. 2011.
- Y. Richter and N. Slonim, "Predicting customer churn in mobile networks through analysis of social groups," *Proceedings of the SIAM International Conference on Data Mining*, pp. 732–741, 2010.