IG-KNN UNTUK PREDIKSI CUSTOMER CHURN TELEKOMUNIKASI

Muhammad Arifin
Fakultas Teknik, Program Studi Sistem Informasi
Universitas Muria Kudus
Email: arifin.m@umk.ac.id

ABSTRAK

IG-KNN merupakan gabungan dari algotitma pemilihan fitur *information gain* dengan algoritma klasifikasi KNN, kedua algoritma ini diharapkan dapat meningkatkan akurasi dalam memprediksi *customer churn* telekomunikasi. Prediksi *customer churn* telekomunikasi merupakan kebutuhan yang sangat penting bagi kelangsungan hidup perusahaan telekomunikasi, dimana dengan banyaknya pelanggang yang meninggalkan perusaan maka perusahaan berpeluang untuk merugi. Mendeteksi pelanggan yang berpeluang meninggalkan perusahaan sejak dini perusahaan akan mendapatkan keuntungan 10 kali, karena biaya untuk mempertahankan pelanggan lebih murah 10 kali lipat dibanding dengan mecari pelanggan baru. Berdasarkan hasil penelitian ini prediksi *customer churn* telekomunikasi dengan menggunakan IG-KNN menunjukkan akurasi yang lebih baik meski dengan nilai k yang berbeda-beda bila dibandingkan dengan prediksi *customer churn* telekomunikasi dengan menggunkan KNN tanpa fitur seleksi *Information Gain*, adapun peningkatan akurasi dari k1 sampai dengan k11 sebesar 1,7%.

Kata kunci: *information gain*, KNN, *customer churn* telekomunikasi.

ABSTRACT

IG-KNN is a combination of information gain feature selection algorithm with KNN classification algorithm, the second algorithm is expected to improve the accuracy of predicting customer churn in telecommunications. Telecommunication customer churn prediction is a very important requirement for the survival of the telecommunications company, where the number of the customers who leave the company for promotional likely to lose money. Detecting corporate customers are likely to leave the company early on will benefit 10 times, because the cost is cheaper to retain customers 10-fold compared to look for new customers. Based on these results the customer churn prediction telecommunications using IG-KNN showed better accuracy, although the value of k different when compared with the prediction of customer churn by using KNN telecommunications without Information Gain feature selection, while the increase in the accuracy of the K1 up to K11 at 1.7%.

Keywords: *information gain, KNN, customer churn telekomunikasi.*

1. PENDAHULUAN

Customer Churn didefinisikan sebagai kecenderungan pelanggan untuk berhenti melakukan bisnis dengan sebuah perusahaan [1]. Hal ini telah menjadi isu penting yang merupakan salah satu tantangan utama oleh banyak perusahaan di era global ini dan harus dihadapinya. [2] mengatakan untuk memperoleh pelanggan baru memerlukan biaya hingga 10 kali lipat lebih mahal dibandingkan biaya untuk mempertahankan pelanggan yang ada. Mahalnya untuk memperoleh pelanggan baru tentunya perusahaan akan lebih memilih mempertahankan pelanggan. Berdasarkan fakta tersebut maka banyak perusahaan sekarang lebih beralih untuk mempertahankan pelanggan dan menghindari *churn* pelanggan.

Prediksi customer churn telekomunikasi menggunakan KNN yang dilakukan [3] menghasilkan nilai akurasi 88% pada nilai K 5 keatas, dengan menggunakan algoritma fitur seleksi diharapkan akan meningkatkan akurasi prediksi selain itu dapat diketahui fitur apa saja yang tidak dibutuhkan dalam memprediksi customer churn telekomunikasi.

Pemilihan Subset fitur, bersama dengan pengaturan parameter dalam prosedur pelatihan *SVM* secara signifikan mempengaruhi akurasi klasifikasi[4]. Jumlah fitur yang banyak (*high dimensional*) akan memberikan masukan yang berbeda-beda. Pada prakteknya, atribut yang berlebihan tidak akan memberikan hasil yang signifikan pada proses pelatihan tetapi justru akan menyebabkan *over-fit* atau *irrelevan* dan *redundan* atribut yang akan menyulitkan algoritma proses pelatihan. Masih sedikit penelitian yang berfokus pada pemilihan fitur untuk mempediksi *customer churn* telekomunikasi[5]. Untuk mencapai kinerja yang baik dan akurasi yang tinggi jumlah fitur yang digunakan untuk

memprediksi *customer churn* cukup enam sampai dengan delapan fitur. Tujuan pemilihan masukan membantu menghapus masukan yang tidak relevan, menghapus masukan yang bergantung dengan masukan lainnya, sehingga pembuatan model lebih ringkas, transparan dan mengurangi waktu untuk pembentukan model[6].

Tujuan seleksi fitur adalah untuk mengidentifikasi beberapa fitur dalam kumpulan data yang sama pentingnya, dan membuang semua fitur lain seperti informasi yang tidak relevan dan berlebihan. Proses seleksi fitur mengurangi dimensi dari data dan memungkinkan algoritma belajar untuk beroperasi lebih cepat dan lebih efektif[7]. Solusi untuk masalah ini adalah langkah *pre-processing* yaitu dengan menghilangkan atribut dari data yang tidak relevan sebelum digunakan pada algoritma *data mining.*

Langkah *pre-processing* yaitu dengan menghilangkan atribut dari data yang tidak relevan sebelum digunakan pada algoritma *data mining* disebut dengan *Feature Selection*, algoritma untuk mengurangi dimensi atribut atau seleksi firut digunakan untuk meningkatkan akurasi dari algoritma klasifikasi, adapun algoritma seleksi fitur yang banyak digunakan diantaranya adalah *Minimum Redudancy and Maximum Relevance* (mRMR)[8–13], *Principle Component Analysis* (PCA)[8], [14], [15], *Fast Correlation Based Feature Selection* (FCBF)[9], [16–20], *Recursive Feature Elimination with SVM* (SVM-RFE)[9], [20–22], *Information Gain* (IG)[10], [14], [18], [20], [22–26] dan lain sebagainya.

Menemukan atribut terbaik adalah hal yang mudah dengan menggunakan *information gain* karena setiap atribut dapat diketahui nilainya dan dapat dipilih yang terbaik[27]. *Information gain* adalah algoritma fitur seleksi yang sangat populer dan berhasil digunakan dalam memilih fitur yang terbaik khususnya dalam menangani data yang berdimensi tinggi[28]. Chizi dan Maimon (2002) menjelaskan beberapa metode baru untuk pemilihan variabel yang didasarkan pada algoritma yang sederhana dan menggunakan *evaluator* terkenal seperti *information gain*, *logistic regression* dan *random selection*. Semua metode disajikan dengan hasil empiris pada dataset dan dengan batas-batas teoritis pada setiap metode[7]. Nilai *information gain* terbesar pada atribut suatu data menunjukkan bahwa atribut tersebut adalah atribut yang paling informatif, yang artinya paling relevan terhadap kelas targetnya. Semakin besar nilai *information gain* pada suatu atribut, maka semakin besar pula pengaruhnya terhadap pengklasifikasian suatu data[29]. Sebuah fitur dengan informasi yang lebih tinggi nilai *gain* menunjukkan diskriminasi lebih tinggi fitur ini dibandingkan dengan kategori lain dan berarti bahwa fitur tersebut berisi informasi gen yang berguna untuk klasifikasi[30].

Pemilihan algoritma seleksi atribut *information gain* pada penelitian ini berdasarkan pada pernyataan-pernyataan peneliti diatas diantaranya bahwa nilai *information gain* terbesar pada atribut suatu data menunjukkan bahwa atribut tersebut adalah atribut yang paling informatif, yang artinya paling relevan terhadap kelas targetnya. Semakin besar nilai *information gain* pada suatu atribut, maka semakin besar pula pengaruhnya terhadap pengklasifikasian suatu data[29], selain itu peneliti lain juga mengungkapkan bahwa *information gain* adalah algoritma fitur seleksi yang sangat populer dan berhasil digunakan dalam memilih fitur yang terbaik khususnya dalam menangani data yang berdimensi tinggi[28]

Pada penelitian ini, akan menggunakan *Information Gain* untuk pemilihan fitur pada tahap *pre-processing* dan menerapkan algoritma *K-NN* sebagai proses pelatihan untuk memprediksi *customer churn* telekomunikasi. Sehingga diharapkan dapat menghasilkan akurasi yang lebih baik.

Prediksi *Customer Churn* saat ini menjadi obyek yang diteliti dalam data mining dan telah diterapkan dalam bidang perbankan, telekomunikasi seluler dan asuransi[31]. Analisa data yang dilakukan secara otomatis dengan menggunakan data mining dan teknologi *machine learning* telah lama diterapkan pada masalah analisis *Customer Churn*. Berdasarkan penelitian pada prediksi *Customer Churn*, Wei dan Chiu[32] mengembangkan sebuah model baru untuk prediksi *Customer Churn* penyedia layanan telekomunikasi dengan menggunakan data mining, pada waktu itu penelitian terakhir menggunakan teknik analisa klasifikasi untuk membangun model prediksi dalam memprediksi *Customer Churn* di industri telekomunikasi.

Beberapa teknik data mining yang populer diusulkan untuk memprediksi *Customer Churn* adalah *K-NN*[2], [8], [33], *Support Vector Machines*[1], [2], [33–35] dan *Logistic Regretion*[2], [33–35].

K-NN sangat mudah dipahami dan sangat efektif untuk klasifikasi [36], Algoritma *K-NN* pertama yang akan dilihat adalah tetangga k-terdekat. Cara kerjanya dapat dilihat sebagai berikut ada sebuah data dimana semua data telah memiliki label dan ketika ada sepotong data baru tanpa label, maka bandingkan potongan baru dengan data yang ada, kemudian ambil potongan-potongan yang paling mirip data (tetangga terdekat) dan melihat label mereka. Terakhir, mengambil suara mayoritas dari potongan k paling mirip data, dan mayoritas adalah kelas baru ditetapkan sebagai hasil klasifikasi. Ketepatan algoritma *K-NN* 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. *K-NN* memiliki beberapa kelebihan, diantaranya adalah ketangguhan terhadap data yang

memiliki banyak noise serta efektif terhadap data yang berukuran sangat besar. Tetapi dibalik itu *K-NN* juga menyimpan kelemahan yang diantaranya adalah perlunya menentukan k secara manual.

Logistic Regression dipresentasikan untuk prediksi dengan menggunakan lebih dari satu *Linear Regression* [37],. *Logistic Regression* menampilkan persamaan linear yang saling berhubungan antara beberapa variabel acak, dimana variabel yang tergantung (*dependent*) adalah variabel yang berkelanjutan. Metode ini merupakan perluasan metode *Linear Regression* yang menggunakan lebih dari satu variabel. Dalam beberapa kasus, variabel yang tergantung menunjuk kepada dua nilai/kategori tidak dapat menggunakan *Linear Regression*, tetapi dapat melakukan pendekatan yang serupa yang dapat disebut juga *Multiple Linear Logistic Regression*. Sedangkan [38] model *Logistic Regression* merupakan probabilitas dari beberapa peristiwa, metode ini menggunakan fungsi linear untuk perhitungan prediksi pada beberapa variabel. Prediksi dengan menggunakan metode *Logistic Regression* dapat digunakan pada variabel yang ditentukan atau varibel yang sudah dikategorikan menjadi 2 variabel. Seperti pada prediksi hidup atau mati, sakit atau tidak sakit, menang atau kalah, maupun *churn* atau tidak seperti yang ada pada data set *customer churn*. Permasalahan umum yang timbul pada *logistic regression* adalah *overfitting* pada data pelatihan, terutama ketika data yang sangat tinggi[39].

Support Vector Machines diperkenalkan oleh Vapnik (1995) menggunakan model linier untuk menerapkan batas kelas *non-linier* dengan masukan *non-linier vektor* ke ruang fitur berdimensi tinggi. Dalam ruang baru, optimal *hyperplane* pemisah dibangun. *SVM* dapat berfungsi sebagai alternative kombinasi yang kuat dari model statistik konvensional[40]. Penggunaan *SVM* untuk memprediksi *customer churn* juga telah mendapat perhatian khusus dalam penelitian terakhir[34] yaitu untuk menyelidiki bagaimana secara efektif *SVM* akan mendeteksi *customer churn*. Kelemahan umum *SVM* adalah kurangnya transparansi hasil dan mempunyai dimensi yang sangat tinggi[41].

Beberapa algoritma klasifikasi diatas memiliki permasalahan dimana permasalahan umum yang timbul pada *logistic regression* adalah *overfitting* pada data pelatihan, terutama ketika data yang sangat tinggi[39], dan kelemahan umum *SVM* adalah kurangnya transparansi hasil dan mempunyai dimensi yang sangat tinggi[41], sedangkan algoritma *K-NN* meskipun handal terhadap data noise namun *K-NN* sangat sensitif terhadap fitur yang tidak relevan atau berlebihan karena semua fitur berkontribusi terhadap klasifikasi. Dalam penelitian yang dilakukan oleh (Verbeke, Dejaegen Martens, Hur dan Baesens 2012) yang meneliti dua puluh satu model prediksi *customer churn* sektor telekomunikasi dari 11 operator dari beberapa negara didunia menyatakan bahwa tidak semua fitur berkontribusi terhadap klasifikasi cukup 6 sampai 8 fitur yang digunakan untuk memprediksi *customer churn*, penelitian ini berfokus dalam melakukan pengurangan fitur pada data *customer churn*, sedangkan untuk membuktikan fitur-fitur yang terpilih adalah fitur yang paling berpengaruh terhadap klasifikasi maka dipilih algoritma klasifikasi *K-NN* dimana algoritma ini sangat sensitif terhadap fitur yang tidak relevan atau berlebihan karena semua fitur berkontribusi terhadap klasifikasi.

Pemilihan Subset fitur, bersama dengan pengaturan parameter dalam prosedur pelatihan *SVM* secara signifikan mempengaruhi akurasi klasifikasi[4]. Jumlah fitur yang banyak (*high dimensional*) akan memberikan masukan yang berbeda-beda. Pada prakteknya, atribut yang berlebihan tidak akan memberikan hasil yang signifikan pada proses pelatihan tetapi justru akan menyebabkan *over-fit* atau *irrelevan* dan *redundan* atribut yang akan menyulitkan algoritma proses pelatihan. Masih sedikit penelitian yang berfokus pada pemilihan fitur untuk mempediksi *customer churn* telekomunikasi[5]. Untuk mencapai kinerja yang baik dan akurasi yang tinggi jumlah fitur yang digunakan untuk memprediksi *customer churn* cukup enam sampai dengan delapan fitur. Tujuan pemilihan masukan membantu menghapus masukan yang tidak relevan, menghapus masukan yang bergantung dengan masukan lainnya, sehingga pembuatan model lebih ringkas, transparan dan mengurangi waktu untuk pembentukan model[6].

2. LANDASAN TEORI

2.1 Data Mining

Data Mining (DM) adalah inti dari proses Knowledge Discovery in Database (KDD), melibatkan menyimpulkan algoritma yang mengeksplorasi data, mengembangkan model dan menemukan pola yang tidak diketahui sebelumnya. Model ini digunakan untuk memahami fenomena dari data, analisis dan prediksi. Aksesibilitas dan banyaknya data membuat *Knowledge Discovery* dan *Data Mining* menjadi masalah yang cukup penting dan dibutuhkan [7]. Tiga proses dalam data mining [37] diantaranya:

1. Menjelajahi data, yang terdiri dari pembersihan data, transformasi data, pengurangan dimensi, pemilihan fitur, dll.

2. Membangun model dan validasi, mengacu pada analisis berbagai model dan memilih satu yang memiliki kinerja terbaik dari evaluasi perkiraan-kompetitif model.
3. Menerapkan model untuk data baru untuk menghasilkan perkiraan yang benar / perkiraan untuk masalah diselidiki.

2.2 Algoritma Klasifikasi

Klasifikasi merupakan salah satu tujuan yang banyak dihasilkan dalam *data mining*. Klasifikasi merupakan proses pengelompokkan sebuah variabel kedalam kelas yang sudah ditentukan [42]. Data mining mampu mengolah data dalam jumlah besar, setiap data terdiri dari kelas tertentu bersama dengan variable dan faktor faktor penentu kelas variabel tersebut. Dengan data mining, peneliti dapat menentukan suatu kelas dari variabel data yang dimiliki.

Ada banyak algoritma klasifikasi yang dapat digunakan dalam *data mining.* Mulai dari *k-nearest neighbor, logistic regresi, neural network, svm* dan lainnya. Dalam penelitian terkait, disimpulkan bahwa algoritma *K-NN* memiliki hasil yang kurang baik untuk proses klasifikasi kelayakan pelanggan dengan atribut yang banyak.

2.3 Pengantar Algoritma Klasifikasi K-NN

K-NN adalah sebuah metode klasifikasi terhadap sekumpulan data berdasarkan pembelajaran data yang sudah terklasifikasikan sebelumya. *K-NN* termasuk dalam golongan *supervised learning*, dimana hasil *query instance* yang baru diklasifikasikan berdasarkan mayoritas kedekatan jarak dari kategori yang ada dalam *K-NN*. Nantinya kelas yang baru dari suatu data akan dipilih berdasarkan grup kelas yang paling dekat jarak vektornya.

Tujuan dari algoritma ini adalah mengklasifikasikan obyek baru berdasarkan atribut dan training *sample*. *Classifier* tidak menggunakan model apapun untuk dicocokkan dan hanya berdasarkan pada memori. Diberikan titik *query*, akan ditemukan sejumlah k obyek atau (titik *training*) yang paling dekat dengan titik *query*. Klasifikasi menggunakan voting terbanyak diantara klasifikasi dari k obyek. Algoritma KNN menggunakan klasifikasi ketetanggaan sebagai nilai prediksi dari *query instance* yang baru.

Algoritma metode *K-NN* sangatlah sederhana [36], bekerja berdasarkan jarak terpendek dari *query instance* ke *training sample* untuk menentukan *K-NN*nya. Training sample diproyeksikan ke ruang 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 klasifikasi yang paling banyak ditemui pada k buah tetangga terdekat dari titik tersebut. Dekat atau jauhnya tetangga biasanya dihitung berdasarkan *Euclidean Distance*.

Ketepatan algoritma *KNN* sangat dipengaruhi oleh ada atau tidaknya fitur-fitur yang tidak *relevant* 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.

K-NN memiliki beberapa kelebihan, diantaranya adalah handal terhadap data yang memiliki banyak *noise* serta efektif terhadap data yang berukuran sangat besar. Tetapi dibalik itu *K-NN* juga menyimpan kelemahan yang diantaranya adalah perlunya menentukan k secara manual dan perlunya menghitung satu persatu data testing terhadap semua data training (tidak ada model yang terbentuk)selain itu *KNN* sangat dipengaruhi oleh ada atau tidaknya fitur-fitur yang tidak *relevant* atau jika fitur tersebut tidak setara dengan relevansinya terhadap klasifikasi.

2.4 Seleksi Fitur

Tujuan dari seleksi fitur adalah untuk mengidentifikasi beberapa fitur dalam data set sama pentingnya, dan membuang fitur lainnya yang memberikan informasi yang tidak *relevant* dan berlebihan (Hall 1999). Proses seleksi fitur mengurangi dimensi dari data dan memungkinkan algoritma klasifikasi untuk beroperasi lebih cepat dan lebih efektif dan dalam beberapa kasus, keakuratan klasifikasi dapat ditingkatkan. Dalam data set terdapat beberapa atribut yang lebih berguna daripada yang lain, beberapa mungkin lebih membantu proses belajar daripada yang lain (Blum dan Langley, 1997). Solusi untuk masalah ini adalah langkah pemrosesan awal yang menghilangkan atribut dari data yang tidak relefan sebelum diterapkan ke sebuah algoritma Data Mining.

2.5 Information Gain

Information gain (IG) dari suatu atribut, diperoleh dari nilai *entropy* sebelum pemisahan dikurangi dengan nilai *entropy* setelah pemisahan[29]. Untuk tujuan pengurangan fitur, pengukuran nilai informasi diterapkan sebagai tahap pra-pengolahan awal. Hanya atribut memenuhi kriteria *(treshold)* yang ditentukan dipertahankan untuk digunakan oleh algoritma klasifikasi[43].

3. METODE PENELITIAN

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. Seleksi fitur
   d. Penerapan algoritma
   e. Evaluasi hasil

4. PEMBAHASAN

4.1 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. Adapaun hal-hal yang dapat diambil dari metode ini diantaranya adalah:

a. Studi Pendahuluan : tahap ini merupakan kegiatan untuk menemukan informasi tentang obyek permasalahan yang ada. Permasalahan-persalahan yang berkembang beberapa tahun terahir dalam sebuah organisasi atau perusahaan khususnya mengenai *customer churn.*
b. 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.
c. 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.

4.2 Metode Eksperimen

Metode eksperimen ini digunakan untuk menganalisa data yaitu memilah label dan variabel yang selanjutnya data digunakan dalam proses prediksi. Adapun tahapan dalam metode ini adalah sebagai berikut:

4.2.1 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.

4.2.2 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
State	untuk 51 negara bagian *District of Columbia*
Account Length	berapa lama akun aktif
Area Code	kode area
Phone Number	nomer telepon yang digunakan sebagai ID pelanggan
International Plan	rencana internasional
Voice Mail Plan	rencana pesan suara
Number Vmail Messages	jumlah pesan *voice mail*
Total Days	total panggilan sehari pada siang hari, yang terdiri dari :
	total day minutes (jumlah layanan per menit),
	total day calls (jumlah panggilan) dan
	total day charge (jumlah biaya)
Total Eve	total panggilan sehari pada sore hari, yang terdiri dari
	total eve minutes (jumlah layanan per menit),
	total eve calls (jumlah panggilan) dan
	total eve charge (jumlah biaya)
Total Night	total panggilan sehari pada malam hari, yang terdiri dari
	total night minutes (jumlah layanan per menit),
	total night calls (jumlah panggilan) dan
	total night charge (jumlah biaya)
Total International	total panggilan yang digunakan untuk panggilan internasional, yang terdiri dari
	total intl minutes (jumlah layanan per menit),
	total intl calls (jumlah panggilan) dan
	total intl charge (jumlah biaya)
Number Customer Service Calls	jumlah panggilan ke layanan pelanggan
Label	indikasi tidak *churn* dan *churn* (*False* and *True*)

4.2.3Seleksi Fitur

Tahap ini merupakan tahap utama dalam penelitian ini dari 21 fitur yang ada nantinya dihitung dan dicari fitur apa saja yang relevan yang berpengaruh terhadap klasifikasi, seleksi fitur menggunakan *information gain* dilakukan dengan cara menghitung nilai gain setiap fitur. Ada 3 tahapan dalam pemilihan fitur menggunakan *Information Gain* diantaranya adalah sebagai berikut:

1. Hitung nilai gain informasi untuk setiap atribut dalam dataset asli.
2. Buang semua atribut yang tidak memenuhi kriteria yang ditentukan.
3. Dataset direvisi.

Pengukuran atribut ini dipelopori oleh Claude Shannon pada teori informasi (1) [44][44], [45] dituliskan sebagai:

$$Info(D) = -\sum_{i=1}^{m} p_i \log_2 (p_i) \quad (1)$$

Keterangan:
D : Himpunan Kasus, m : Jumlah partisi D, pi : Proporsi dari Di terhadap D
Dalam hal ini pi adalah probabilitas sebuah *tuple* pada D masuk ke kelas C_i dan diestimasi dengan $|C_i,D|/|D|$. Fungsi log diambil berbasis 2 karena informasi dikodekan berbasis bit.

Selanjutnya mencari nilai entropy setelah pemisahan dengan cara sebagai berikut:

$$Info_A(D) = -\sum_{j=1}^{v} \frac{|D_j|}{|D|} \times I(D_j) \quad (2)$$

Keterangan:
D : Himpunan kasus
A : Atribut
v : Jumlah partisi atribut A
|Dj| : Jumlah kasus pada partisi ke j

|D| : Jumlah kasus dalam D
$I(D_j)$: total entropy dalam partisi

Untuk mencari nilai *Information Gain* atribut A diperoleh dengan persamaan berikut:

$$Gain(A) = I(D) - I(A) \quad (3)$$

Keterangan:
Gain(A) : information atribut A
I(D) : total entropy
I(A) : entropy A

Dengan penjelasan lain, Gain(A) adalah reduksi yang diharapkan di dalam entropi yang disebabkan oleh pengenalan nilai atribut dari A. Atribut yang memiliki nilai *information gain* terbesar dipilih sebagai uji atribut untuk himpunan S. Selanjutnya suatu simpul dibuat dan diberi label dengan label atribut tersebut, dan cabang-cabang dibuat untuk masing-masing nilai dari atribut.

Berdasarkan perhitungan nilai *information gain* tiap atribut menunjukkan bahwa atribut *area code* dan *total night calls* bernilai terendah yaitu 0,000 sedangkan *account length, total day calls, total eve calls, total night minutes,* dan *total night charge* bernilai 0,001 atribut *total intl calls* bernilai 0,005 adapun atribut *total eve minutes, total eve charge, total int minutes* dan *total int charge* bernilai 0,006 adapun atribut *voice mail plane* dan *number vmail messages* bernilai 0,010 atribut *state* dan *international plan* masing-masing memiliki nilai 0,014 dan 0,036 selanjutnya adalah *number customer service calls* bernilai 0,048 dan nilai atribut tertinggi dari data *customer churn* ini adalah atribut *total day minutes* dan *total day charge* yaitu dengan nilai gain 0,056. Setelah diketahui nilai gain dari setiap atributnya langkah selanjutnya adalah menentukan nilai ambang, dan dalam penelitian ini nilai ambang yang diambil adalah 0,005 sehingga atribut yang terpilih adalah 12 atribut diantaranya *total intl calls, total eve minutes, total eve charge, total int minutes, total int charge, voice mail plane, number vmail messages, state, international plan, number customer service calls, total day minutes* dan *total day charge*

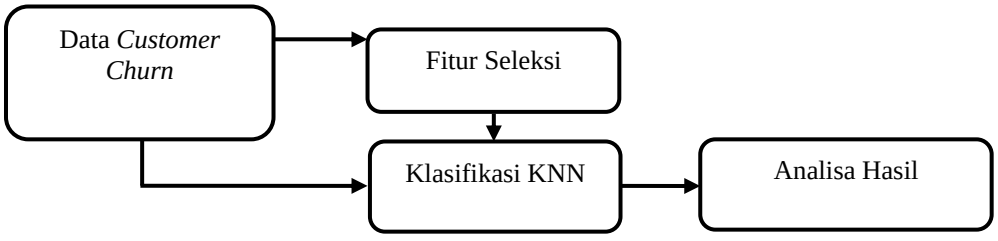
4.2.4 Penerapan Algoritma

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. 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 4.

$$D(a,b) = \sqrt{\sum_{k=1}^{d}(a_k - b_k)^2} \quad (4)$$

Keterangan:
D(a,b) : jarak skalar dari dua buah vektor a dan b dari matrik berukuran D dimensi
k : data traning ke n
d : jumlah data training
a : data training
b : data testing

Persamaan 4 adalah persamaan untuk mencari jarak terpendek antara data testing dengan data taining. Untuk mengetahui proses lengkap dari prediksi *customer churn* telekomunikasi ditunjukkan pada gambar 1.



Gambar 1. Blok Diagram Algoritma Prediksi *K-NN*

Gambar 1 blok diagram algoritma prediksi *K-NN* diatas memperlihatkan langkah-langkah dalam memprediksi data *chustomer churn*, data *chustomer churn* diklasifikasi menggunakan algoritma KNN dibandingkan dengan data *chustomer churn* yang fiturnya telah diseleksi menggunakan *information gain* sebelum diklasifikasi.

4.2.5 Evaluasai Hasil

Dari berbagai eksperimen yang dilakukan dalam memprediksi *customer churn* Telekomunikasi dengan menggunakan Algoritma KNN diperoleh data pada Tabel 2 :

Tabel 2. Hasil Prediksi Customer Churn Telekomunikasi

Prediksi	K											
	1		3		5		7		9		11	
	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
IG-KNN	**85,1**	**0,5**	**89**	**0,7**	**89,4**	**0,71**	**89,6**	**0,73**	**89,7**	**0,73**	**89,8**	**0,73**

5. KESIMPULAN

Berdasarkan hasil eksperimen, mulai tahap awal hingga evaluasi, dapat ditarik kesimpulan bahwa model prediksi *customer churn* menggunakan metode K-NN dengan pengurangan fitur menggunakan *information gain* cukup akurat dibandingkan dengan tanpa menggunakan fitur seleksi dimana dengan nilai k yang berbeda-beda algoritma IG-KNN tetap menunjukkan hasil yang paling baik. Peningkatan akurasi dari k1 sampai dengan k11 sebesar 1,7%.

6. SARAN

Penelitian ini telah menghasilkan suatu model prediksi yang optimal dan akurat, namun untuk penelitian selanjutnya masih memerlukan pengembangan dalam beberapa hal, yakni:

1. Diperlukan simulasi prediksi dengan jumlah data yang lebih banyak, sehingga analisa akan bertambah optimal dan akurat
2. Bandingkan algoritma fitur seleksi lain untuk mendapatkan hasil yang lebih baik.
3. Membuat aplikasi sistem/aplikasi *online* dengan menerapkan metode lain untuk prediksi *customer churn* lebih dinamis.

7. DAFTAR PUSTAKA

[1] 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.

[2] 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.

[3] M. Arifin, "BUSINESS INTELLIGENCE UNTUK PREDIKSI CUSTOMER CHURN TELEKOMUNIKASI," in *SEMINAR NASIONAL TEKNOLOGI DAN INFORMATIKA*, 2014, pp. 279–286.

[4] M. Zhao, C. Fu, L. Ji, K. Tang, and M. Zhou, "Feature selection and parameter optimization for support vector machines: A new approach based on genetic algorithm with feature chromosomes," *Expert Systems with Applications*, vol. 38, no. 5, pp. 5197–5204, May 2011.

[5] W. Verbeke, K. Dejaeger, D. Martens, J. Hur, and B. Baesens, "New insights into churn prediction in the telecommunication sector: A profit driven data mining approach," *European Journal of Operational Research*, vol. 218, no. 1, pp. 211–229, Apr. 2012.

[6] Y. Zhang, R. Liang, Y. Li, Y. Zheng, and M. Berry, "Behavior-Based Telecommunication Churn Prediction with Neural Network Approach," *2011 International Symposium on Computer Science and Society*, pp. 307–310, Jul. 2011.

[7] O. Maimon and L. Rokach, *Data Mining and Knowledge Discovery Handbook*, Second. Boston, MA: Springer US, 2010, pp. 1, 86, 97.

[8] A. Idris, M. Rizwan, and A. Khan, "Churn prediction in telecom using Random Forest and PSO based data balancing in combination with various feature selection strategies," *Computers & Electrical Engineering*, vol. 38, no. 6, pp. 1808–1819, Nov. 2012.

[9] A. J. Ferreira and M. A. T. Figueiredo, "Efficient feature selection filters for high-dimensional data," *Pattern Recognition Letters*, vol. 33, no. 13, pp. 1794–1804, Oct. 2012.

[10] B. Chandra and M. Gupta, "An efficient statistical feature selection approach for classification of gene expression data.," *Journal of biomedical informatics*, vol. 44, no. 4, pp. 529–35, Aug. 2011.

[11] C. Ding and H. Peng, "Minimum Redundancy Feature Selection from Microarray Gene Expression Data," *Proceedings of the Computational Systems Bioinformatics*, pp. 0–5, 2003.

[12] C. Ding and H. Peng, "Minimum redundancy feature selection from microarray gene expression data," *Journal of Bioinformatics and Computational Biology*, vol. 3, no. 2, pp. 185–205, 2005.

[13] H. Peng, F. Long, and C. Ding, "Feature Selection Based on Mutual Information : Criteria of Max-Dependency, Max-Relevance, and Min-Redundancy," *IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE*, vol. 27, no. 8, pp. 1226–1238, 2005.

[14] H. Uğuz, "A hybrid system based on information gain and principal component analysis for the classification of transcranial Doppler signals.," *Computer methods and programs in biomedicine*, vol. 107, no. 3, pp. 598–609, Sep. 2012.

[15] C. Tsai and Y. Hsiao, "Combining multiple feature selection methods for stock prediction : Union , intersection , and multi-intersection approaches," *Decision Support Systems*, vol. 50, no. 1, pp. 258–269, 2010.

[16] B. Senliol, G. Gulgezen, L. Yu, and Z. Cataltepe, "Fast Correlation Based Filter (FCBF) with a different search strategy," in *2008 23rd International Symposium on Computer and Information Sciences*, 2008, pp. 1–4.

[17] M. Blachnik, W. Duch, A. Kachel, and J. Biesiada, "Feature Selection for Supervised Classification : A Kolmogorov- Smirnov Class Correlation-Based Filter," *Blachnik09featureselection*, 2009.

[18] T. Deepa and L. Ladha, "Feature Selection Methods And Algorithms," *International Journal on Computer Science and Engineering (IJCSE)*, vol. 3, no. 5, pp. 1787–1797, 2011.

[19] L. Yu and H. Liu, "Feature Selection for High-Dimensional Data : A Fast Correlation-Based Filter Solution," *Proceedings of the Twentieth International Conference on Machine Learning (ICML-2003)*, 2003.

[20] H. Deng and G. Runger, "Feature Selection via Regularized Trees," *International Conference Neural Network*, 2012.

[21] C. Kang, "Customer Churn Prediction Based on SVM-RFE," *2008 International Seminar on Business and Information Management*, pp. 306–309, Dec. 2008.

[22] F. Tan, "Improving Feature Selection Techniques for Machine Learning," *Dissertation Georgia State University*, 2007.

[23] C. Lee and G. G. Lee, "Information gain and divergence-based feature selection for machine learning-based text categorization," *Information Processing & Management*, vol. 42, no. 1, pp. 155–165, Jan. 2006.

[24] J. Dai and Q. Xu, "Attribute selection based on information gain ratio in fuzzy rough set theory with application to tumor classification," *Applied Soft Computing*, vol. 13, no. 1, pp. 211–221, Jan. 2013.

[25] B. Azhagusundari and A. S. Thanamani, "Feature Selection based on Information Gain," *International Journal of Innovative Technology and Exploring Engineering (IJITEE)*, vol. 2, no. 2, pp. 18–21, 2013.

[26] J. Novakovic, "The Impact of Feature Selection on the Accuracy of 1DwYH Bayes Classifier," vol. 2, pp. 1113–1116, 2010.

[27] C. Aggarwal and P. S. Yu, *Privacy preserving data mining Models and Algorithms*. New York: Springer US, 2008, p. 360.

[28] I. Koprinska, "Feature Selection for Brain-Computer Interfaces," *PAKDD Workshops*, pp. 100–111, 2010.

[29] W. Maharani, "Klasifikasi Data Menggunakan JST Backpropagation," *Seminar Nasional Informatika 2009 (semnasIF 2009)*, vol. 2009, no. semnasIF, pp. 25–31, 2009.

[30] C. Yang and L. Chuang, "IG-GA : A Hybrid Filter / Wrapper Method for Feature Selection of Microarray Data," *Journal of Medical and Biological Engineering*, vol. 30, no. 1, pp. 23–28, 2009.

[31] L. J. S. . Alberts, *Churn Prediction in The Mobile Telecommunications Industry*, Thesis., no. September. Maastricht: Maastricht University, 2006, p. 6.

[32] A. T. Jahromi, "MASTER ' S THESIS Predicting Customer Churn in Telecommunications Service Providers," 2009.

[33] Y. Huang and T. Kechadi, "An effective hybrid learning system for telecommunication churn prediction," *Expert Systems with Applications*, no. May, May 2013.

[34] K. Coussement and D. Van den Poel, "Churn prediction in subscription services: An application of support vector machines while comparing two parameter-selection techniques," *Expert Systems with Applications*, vol. 34, no. 1, pp. 313–327, Jan. 2008.

[35] B. Huang, M. T. Kechadi, and B. Buckley, "Customer churn prediction in telecommunications," *Expert Systems with Applications*, vol. 39, no. 1, pp. 1414–1425, Jan. 2012.

[36] P. Harrington, *Machine Learning in Action*. USA: Manning Publications, 2012, p. 18.

[37] F. Gorunescu, *Data Mining Concepts,Models and Techniques*, vol. 12. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, p. 7 116 120.

[38] J. Han and M. Kamber, *Data Mining: Concepts and Techniques*, Second. United States of America: Morgan Kaufmann publications, 2007, p. 358.

[39] T. M. Mitchell, "Generative And Discriminative Classifiers : Naive Bayes And Logistic Regression," in *Machine Learning*, 2010, pp. 1–17.

[40] P. Ravisankar, V. Ravi, G. Raghava Rao, and I. Bose, "Detection of financial statement fraud and feature selection using data mining techniques," *Decision Support Systems*, vol. 50, no. 2, pp. 491–500, Jan. 2011.

[41] L. Auria and R. A. Moro, "Support Vector Machines (SVM) as a Technique for Solvency Analysis," *SSRN Electronic Journal*, no. August, 2008.

[42] D. T. Larose, *DISCOVERING KNOWLEDGE IN DATA An Introduction to Data Mining*. 2005, p. 14.

[43] M. Bramer, *Principles of Data Mining*. London: Springer US, 2007, pp. 80, 154.

[44] R. G. Gallager and L. Fellow, "Claude E . Shannon : A Retrospective on His Life , Work , and Impact," *IEEE TRANSACTIONS ON INFORMATION THEORY*, vol. 47, no. 7, pp. 2681–2695, 2001.

[45] K. M. Risvik, *Discretization of Numerical Attributes Preprocessing for Machine Learning*. Trondheim, Norway: Department of Computer and Information Science Norwegian University of Science and Technology, 1997, p. 29.