

# Penerapan Bootstrapping untuk Ketidakseimbangan Kelas dan *Weighted* Information Gain untuk *Feature Selection* pada Algoritma Support Vector Machine untuk Prediksi Loyalitas Pelanggan

Abdul Razak Naufal, Romi Satria Wahono dan Abdul Syukur

Fakultas Ilmu Komputer, Universitas Dian Nuswantoro

*abdul.razak.naufal@gmail.com, romi@romisatriawahono.net, abdul\_s@dosen.dinus.ac.id*

Abstrak: Prediksi loyalitas pelanggan merupakan sebuah strategi bisnis yang penting bagi industri telekomunikasi modern untuk memenangkan persaingan global, karena untuk mendapatkan pelanggan baru biayanya lebih mahal lima sampai enam kali lipat daripada mempertahankan pelanggan yang sudah ada. Klasifikasi loyalitas pelanggan bertujuan untuk mengidentifikasi pelanggan yang cenderung beralih ke perusahaan kompetitor yang sering disebut *customer churn*. Algoritma Support Vector Machine (SVM) adalah algoritma klasifikasi yang juga berfungsi untuk memprediksi loyalitas pelanggan. Penerapan algoritma SVM dalam memprediksi loyalitas pelanggan mempunyai kelemahan yang mempengaruhi keakuratan dalam memprediksi loyalitas pelanggan yaitu sulitnya pemilihan fungsi kernel dan penentuan nilai parameternya. Dataset yang besar pada umumnya mengandung ketidakseimbangan kelas (*class imbalance*), yaitu adanya perbedaan yang signifikan antar jumlah kelas, yang mana kelas negatif lebih besar daripada kelas positif. Dalam penelitian ini diusulkan metode *resampling* bootstrapping untuk mengatasi ketidakseimbangan kelas. Selain itu dataset juga mengandung fitur yang tidak relevan sehingga dalam pemilihan fitur dalam penelitian ini digunakan metode dua fitur seleksi yaitu Forward Selection (FS) dan *Weighted* Information Gain (WIG). FS berfungsi untuk menghilangkan fitur yang paling tidak relevan serta membutuhkan waktu komputasi yang relatif pendek dibandingkan dengan backward elimination dan stepwise selection. WIG digunakan untuk memberi nilai bobot pada setiap atribut, karena WIG lebih cocok digunakan dalam memilih fitur terbaik daripada Principal Component Analysis (PCA) yang biasa digunakan untuk mereduksi data yang berdimensi tinggi. Tujuan pembobotan ini untuk meranking atribut yang memenuhi kriteria (*threshold*) yang ditentukan dipertahankan untuk digunakan oleh algoritma SVM. Sedangkan untuk pemilihan parameter algoritma SVM dengan menggunakan metode *grid search*. Metode *grid search* dapat mencari nilai parameter terbaik dengan memberi *range* nilai parameter. *Grid search* juga sangat handal jika diaplikasikan pada dataset yang mempunyai atribut sedikit daripada menggunakan *random search*. Hasil eksperimen dari beberapa kombinasi parameter dapat disimpulkan bahwa prediksi loyalitas pelanggan dengan menggunakan sampel bootstrapping, FS-WIG serta *grid search* lebih akurat dibanding dengan metode individual SVM.

**Kata Kunci:** loyalitas pelanggan, bootstrapping, *weighted* information gain, support vector machine

## 1 PENDAHULUAN

Industri telekomunikasi merupakan salah satu industri teknologi tinggi yang berkembang paling cepat diantara industri disektor lainnya. Terbukanya persaingan bebas diperusahaan jasa telekomunikasi juga merupakan salah satu

tantangan serius yang harus dihadapi oleh industri telekomunikasi (Huang, Kechadi, & Buckley, 2012), dimana dengan banyaknya kompetitor dibidang telekomunikasi, memaksa perusahaan telekomunikasi harus lebih meningkatkan pelayanan terhadap pelanggan agar tidak pindah ke operator lain (Jadhav & Pawar, 2011), karena biaya untuk mendapatkan pelanggan baru lebih mahal daripada mempertahankan pelanggan yang sudah ada. Dengan *market share* yang besar maka perlu dilakukan usaha-usaha untuk mempertahankan pelanggan agar *market share* yang telah diraih tidak menurun, sehingga perlu menerapkan suatu strategi yang handal dengan mengeluarkan sedikit biaya terhadap memprediksi loyalitas pelanggan namun hasilnya yang besar bisa diraih. Karena untuk mendapatkan pelanggan baru biayanya lebih mahal lima sampai enam kali lipat daripada mempertahankan pelanggan yang sudah ada (Verbeke, Dejaeger, Martens, Hur, & Baesens, 2012). Pelanggan yang meninggalkan atau berhenti terhadap layanan yang telah diberikan oleh perusahaan telekomunikasi dan menjadi pelanggan perusahaan komunikasi kompetitor disebut dengan perilaku *customer churn* (Yu, Guo, Guo, & Huang, 2011), perilaku ini menjadi salah satu kerugian pendapatan perusahaan. Hal ini juga telah menjadi isu penting dan merupakan salah satu tantangan utama perusahaan yang harus dihadapi di era global ini.

Tiap bulan rata-rata pelanggan yang pindah ke operator lain diperusahaan telekomunikasi di Eropa mencapai 8%-12% dan biaya perpindahan pelanggan sangat besar yaitu mencapai sekitar 500 EURO (Richeldi & Perrucci, 2002). Sementara di Amerika Serikat, per bulan *churn rate* pada tahun 1998 adalah 2% sampai 3%. Biaya yang harus dikeluarkan untuk memperoleh seorang pelanggan baru rata-rata mencapai USD 400 dan biaya perpindahan pelanggan pada industri telekomunikasi ini mendekati USD 6,3 juta. Total kerugian per tahun dapat meningkat menjadi USD 9,6 juta apabila kerugian pendapatan per bulan dari pelanggan juga ikut diperhitungkan, sehingga pada suatu perusahaan yang memiliki 1,5 juta pelanggan, pengurangan loyalitas pelanggan dari 2% menjadi 1% dapat meningkatkan pendapatan tahunan sedikitnya sebanyak USD 54 juta dolar (Mozer, Wolniewicz, Grimes, Johnson, & Kaushansky, 2000). Dari data dan fakta inilah bahwa mengatasi perilaku loyalitas pelanggan merupakan suatu bisnis proses yang dapat meningkatkan keuntungan dan bisa menjadi salah satu kunci sukses untuk memenangkan persaingan di dunia bisnis jasa telekomunikasi. Telekomunikasi merupakan salah satu industri, dimana basis pelanggan memainkan peran penting dalam mempertahankan pendapatan yang stabil, dengan demikian perhatian serius dikhususkan untuk mempertahankan pelanggan sehingga menggeser fokus utama perusahaan dari mencari *market share* menjadi mempertahankan pelanggan.

Prediksi loyalitas pelanggan merupakan sebuah strategi bisnis yang sangat penting bagi perusahaan telekomunikasi

modern. Pendekatan mendasar yang dilakukan untuk memprediksi loyalitas pelanggan dapat diperoleh dari kemungkinan loyalitas pelanggan dengan menggunakan metode prediksi. Prediksi loyalitas pelanggan bertujuan untuk mengidentifikasi pelanggan yang cenderung beralih ke perusahaan kompetitor (Z.-Y. Chen, Fan, & Sun, 2012). Oleh karena itu, banyak perusahaan bisnis menggunakan prediksi loyalitas pelanggan untuk mengidentifikasi pelanggan yang cenderung berpindah operator. Tindakan prediksi loyalitas pelanggan ini dapat diambil untuk membantu perusahaan dalam meningkatkan strategi intervensi untuk meyakinkan pelanggan supaya tetap berlangganan dan mencegah hilangnya bisnis perusahaan (Z.-Y. Chen et al., 2012). Hal ini memungkinkan pengelolaan pelanggan yang efisien dan alokasi yang lebih baik dari sumber daya marketing untuk kampanye dalam mempertahankan pelanggan.

Serangkaian proses untuk mendapatkan pengetahuan atau pola dari sekumpulan data disebut dengan data mining (Witten, I. H., Frank, E., & Hall, 2011). Klasifikasi merupakan topik penelitian yang penting dalam data mining, karena teknik klasifikasi dapat memecahkan masalah dengan menganalisis data dalam database yang masing-masing data dikelompokkan ke dalam kelas tertentu. Ukuran data yang sangat besar pada perusahaan telekomunikasi (Farvaresh & Sepehri, 2011) menjadi rintangan utama dalam mencapai kinerja yang diinginkan untuk memprediksi loyalitas pelanggan.

Ada beberapa teknik data mining untuk memprediksi loyalitas pelanggan yang telah dilakukan oleh para peneliti yaitu (Huang et al., 2012) dengan menggunakan algoritma Neural Network (NN) (Tsai & Lu, 2009), Decision Tree (DT) (Han, Lu, & Leung, 2012), Logistic Regression (LR) (Nie, Rowe, Zhang, Tian, & Shi, 2011), Support Vector Machine (SVM) (Xia & Jin, 2008). Dari beberapa penelitian yang telah dilakukan tersebut dapat disimpulkan bahwa neural network mempunyai kemampuan dalam memprediksi hasil keputusan diagnostik dibandingkan dengan algoritma logistic regression, NN juga mempunyai kemampuan dalam mendeteksi hubungan kompleks yang bersifat *nonlinear* antara faktor prediksi dan hasil prediksi (H. Chen, Zhang, Xu, Chen, & Zhang, 2012), tetapi NN mempunyai kelemahan pada perlunya data training yang besar, sulit mengenali pola apabila menggunakan data yang berdimensi tinggi sehingga memerlukan waktu komputasi yang lama (Pan, Iplikci, Warwick, & Aziz, 2012) dan sering mengalami *over-fitting* (Rynkiewicz, 2012). Algoritma SVM bisa mengatasi kelemahan pada NN yaitu masalah tidak membutuhkan data training yang besar (Vapnik, 1998) serta memiliki kemampuan generalisasi yang baik ketika diterapkan pada data noise (Farvaresh & Sepehri, 2011) yang secara teori lebih memuaskan daripada metode neural network (Nugroho, 2008). Selain itu algoritma klasifikasi SVM juga mempunyai resiko yang kecil terhadap masalah *over-fitting* dibanding dengan metode lainnya (Han, J., & Kamber, 2012), tetapi algoritma support vector machine mempunyai kelemahan pada sulitnya pemilihan parameter dan fungsi kernel yang optimal untuk mendapatkan pendekatan yang lebih baik yang secara signifikan dapat mempengaruhi akurasi (Wu, Xindong & Kumar, 2009) (Coussement & Van den Poel, 2008) (Xia & Jin, 2008) (Z. Chen & Fan, 2013) (Wu, 2011). Jadi pemilihan parameter dan fungsi kernel yang tepat sangat mempengaruhi hasil prediksi dalam memprediksi loyalitas pelanggan.

Algoritma Support Vector Machine (SVM) adalah algoritma klasifikasi dalam data mining yang berfungsi untuk memprediksi loyalitas pelanggan, tetapi penerapan algoritma support vector machine dalam memprediksi loyalitas

pelanggan mempunyai kelemahan yang mempengaruhi keakuratan dalam memprediksi loyalitas pelanggan yaitu sulitnya pemilihan fungsi kernel dan penentuan nilai parameter yang tepat. Dataset yang besar pada umumnya mengandung ketidakseimbangan kelas (*class imbalance*), yaitu adanya perbedaan yang signifikan antar jumlah kelas, yang mana kelas negatif lebih besar daripada kelas positif. Dalam penelitian ini diusulkan metode *resampling* bootstrapping untuk mengatasi ketidakseimbangan kelas. Selain itu dataset juga mengandung fitur yang tidak relevan sehingga dalam pemilihan fitur dalam penelitian ini digunakan metode dua fitur seleksi yaitu Forward Selection (FS) dan *Weighted Information Gain* (WIG). FS berfungsi untuk menghilangkan fitur yang paling tidak relevan serta membutuhkan waktu komputasi yang relatif. WIG digunakan untuk memberi nilai bobot pada setiap atribut. Tujuan pembobotan ini untuk merangking atribut yang memenuhi kriteria (*threshold*) yang ditentukan dipertahankan untuk digunakan oleh algoritma SVM. Sedangkan untuk pemilihan parameter algoritma SVM dengan menggunakan metode *grid search*. Metode *grid search* dapat mencari nilai parameter terbaik dengan memberi *range* nilai parameter. Sedangkan fungsi kernel yang akan digunakan dalam penelitian ini yaitu dengan menggunakan fungsi kernel Radial Basis Function (RBF).

## 2 LANDASAN TEORI

### 2.1 PENELITIAN TERKAIT

Beberapa peneliti menggunakan metode untuk mengatasi sulitnya pemilihan parameter dan fungsi kernel pada algoritma support vector machine diantaranya yang dilakukan oleh Kristof Coussement dan Dirk Van den Poel (Coussement & Van den Poel, 2008), untuk mengatasi sulitnya pemilihan parameter dan fungsi kernel menggunakan fungsi kernel Radial Basis Function (RBF) dengan dua parameter yang bermanfaat mengurangi sulitnya membaca data numerik karena nilai kernel terletak antara nol dan satu. Penelitian yang dilakukan oleh Qi Wu (Wu, 2011), diusulkan metode Adaptive and Couchy Mutation Particle Swarm Optimization (ACPSO) sebagai optimasi parameter SVM, karena ACPSO dianggap sebagai teknik yang baik untuk memecahkan masalah kombinatorial, selain itu juga dapat menentukan *hyperparameter* secara bersamaan dan mempunyai kemampuan pencarian global yang kuat, tetapi ACPSO diyakini memiliki ketergantungan yang sensitif pada parameter dan cenderung terjebak dalam local optimum. Selanjutnya penelitian yang dilakukan oleh Chen dan Fan (Z. Chen & Fan, 2013). menggunakan Multi Kernel Support Vector Regression (MK-SVR) yaitu dengan mengoptimalkan kernel dengan mengkombinasikan beberapa kernel dasar yang masing-masing memiliki *hyperparameter* yang identik, multi kernel ini diformulasikan dengan menerapkan strategi dua iterasi secara berulang-ulang yang bermanfaat untuk mempelajari *lagrange multipliers* oleh *Quadratic Programming* (QP) dan pembobotan koefisien dari dua kernel oleh *linear program* sehingga bisa mengontrol variabel independen. Hasil penelitian terkait ini menunjukkan bahwa penggunaan fungsi kernel dan parameter yang tepat dapat meningkatkan performa algoritma support vector machine.

Ada beberapa metode untuk mengatasi masalah ketidakseimbangan kelas (*class imbalance*) salah satunya adalah dengan teknik *resampling* (Wahono, Herman, & Ahmad, 2014) (Yap et al., 2014) (Farvaresh & Sepehri, 2011) (Burez & Van den Poel, 2009) yang dikolaborasikan dengan seleksi fitur (Khoshgoftaar & Gao, 2009) (Lin, Ravitz, Shyu,

& Chen, 2008), oleh karena itu untuk mengatasi masalah *class imbalance* akan digunakan metode sampel bootstrapping. Setelah mendapatkan data *sampling* hasil bootstrapping, kemudian di filter kembali menggunakan metode forward selection. Forward selection sebagai teknik seleksi fitur yang juga berperan dalam mengangani data yang berdimensi tinggi serta mengandung ketidakseimbangan kelas (Maldonado, Weber, & Famili, 2014) (Idris, Rizwan, & Khan, 2012) dengan memilih subset yang tepat dari set fitur asli, karena tidak semua fitur relevan dengan masalah, bahkan ada beberapa dari fitur tersebut malah menjadi penghalang yang dapat mengurangi akurasi. Pemilihan pembobotan atribut pada penelitian ini dengan menggunakan *Weighted Information Gain* (WIG), karena dengan menggunakan WIG setiap atribut dapat diketahui nilainya dan dapat dipilih yang terbaik (Charu C. Aggarwal, 2008), selain itu WIG merupakan algoritma yang sangat cocok digunakan dalam memilih fitur yang terbaik khususnya dalam merangking data (Novakovic, 2010). Dan untuk pemilihan parameter pada algoritma SVM dengan menggunakan metode *grid search*. Metode *grid search* digunakan karena sangat handal jika diaplikasikan pada dataset yang mempunyai atribut sedikit (Bergstra & Bengio, 2012) daripada metode *random search*.

Pemilihan fungsi kernel yang tepat adalah hal yang sangat penting, karena fungsi kernel ini akan menentukan *feature space* dimana fungsi klasifier akan dicari. Dalam penelitian ini akan diusulkan metode bootstrapping, FS-WIG untuk seleksi fitur dan *grid search* sebagai pemilihan parameter pada algoritma support vector machine. Dalam penelitian ini digunakan fungsi kernel Radial Basis Function (RBF), karena kernel RBF menunjukkan *tradeoff* parameter C dalam algoritma support vector machine yang sangat mempengaruhi hasilnya (Zhou, Liu, & Ye, 2009), beberapa eksperimen juga menunjukkan bahwa kernel RBF menghasilkan tingkat kesalahan klasifikasi yang kecil serta mempercepat perhitungan komputasinya (Xu et al., 2014), hal ini sangat cocok dengan sifat dataset yang besar sebagaimana dataset yang digunakan dalam penelitian ini.

## 2.2 TINJAUAN PUSTAKA

### 2.2.1 VARIABEL PREDIKSI LOYALITAS PELANGGAN

Loyalitas pelanggan banyak diteliti dan diterapkan oleh perusahaan, karena dengan mengetahui loyalitas pelanggan perusahaan dapat lebih meningkatkan relasi terhadap pelanggan yang loyal maupun yang tidak loyal, selain itu perusahaan juga dapat menerapkan strategi bisnis untuk memenangkan persaingan dengan perusahaan kompetitor. Ada beberapa faktor dalam menilai loyalitas pelanggan, termasuk menganalisa klasifikasi terhadap *lifecycle* pelanggan. Pelanggan dibagi menjadi 2 jenis, yaitu pelanggan jenis pegawai dan individu. Pelanggan jenis pegawai terdiri dari pegawai asuransi, lembaga pemerintah, angkatan bersenjata dari pemerintah, transportasi dan logistik, departemen energi, pendidikan, pariwisata, hotel dan restoran, bar internet, rumah sakit, bank, operator telekomunikasi dan ISP, agen partai dan organisasi sosial, industri manufaktur, perusahaan besar, menengah dan usaha kecil). Sedangkan pelanggan jenis individu terdiri dari pelanggan pribadi dan pelanggan yang tidak jelas seperti pelanggan yang tinggal dikota dan pelanggan yang tinggal disuatu Negara dalam jangka waktu tertentu (Han et al., 2012). Selain itu faktor dari segi waktu lamanya menjadi pelanggan dibagi menjadi 3 kelompok. Kelompok pertama yang menjadi pelanggan lebih dari 36 bulan tetapi tidak lebih dari 60 bulan, kelompok kedua yang menjadi pelanggan 18 bulan tetapi tidak lebih dari 36 bulan, kelompok ketiga yang

menjadi pelanggan kurang dari 18 bulan (Han et al., 2012). Jadi dalam penelitian ini beberapa faktor utama yang menjadi penilaian atas loyalitasnya pelanggan terdiri dari (Huang et al., 2012):

#### a. *Demographic profiles*

Menggambarkan pengelompokan demografis atau segmen pasar dan informasi demografis mengandung kemungkinan perilaku pelanggan. Biasanya, informasi ini meliputi usia, pekerjaan, jenis kelamin, dan lain-lain.

#### b. *Information of grants*

Beberapa pelanggan telah memperoleh beberapa hibah khusus sehingga tagihan mereka dibayar penuh atau sebagian oleh pihak ketiga. Sebagai contoh, pelanggan yang cacat atau usia lebih dari 80 tahun yang ingin melanjutkan layanan.

#### c. *Customer account information*

Informasi ini berisi jenis paket layanan, indikator pengontrol kredit, indikator junk mail, tanggal pertama menggunakan layanan, tanggal pembuatan, frekuensi tagihan, saldo rekening, sewa peralatan, jenis pembayaran dan atribut durasi panggilan, jumlah panggilan dan harga standar dan biaya.

#### d. *The historical information of bills and payments*

Informasi ini menyangkut penagihan biaya setiap pelanggan dan jumlah layanan langganan pelanggan dalam setahun.

#### e. *Complaint Information*

Keluhan pelanggan merupakan masalah yang terjadi yang merugikan pelanggan kemudian disampaikan kepada perusahaan. Pada perusahaan yang diteliti keluhan yang dicatat adalah keluhan yang disampaikan secara langsung oleh perusahaan ataupun keluhan yang datang saat pelanggan dikunjungi oleh marketing.

#### f. *Call details*

Pada perusahaan jasa telekomunikasi rincian panggilan mengacu pada durasi panggilan, harga dan jenis panggilan, misalnya seberapa sering pengguna telepon lokal, interlokal, internasional, atau apakah pelanggan berlangganan juga terhadap produk internet atau yang telah ditawarkan lainnya, dan pelanggan yang jumlah pemakaian teleponnya sedikit, layanan telepon sedikit seperti sms mereka inilah yang masuk ke kelompok pelanggan yang *churn*.

#### g. *Incoming calls details*

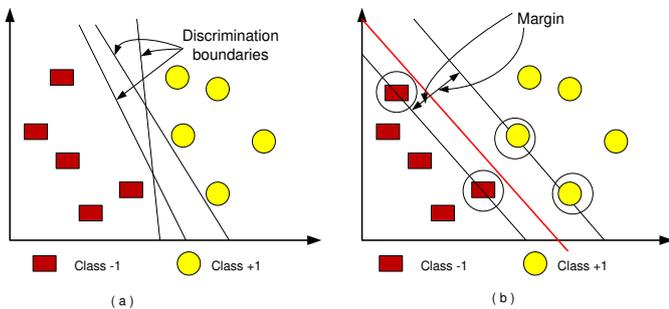
Rincian panggilan yang diterima termasuk durasi panggilan dan jumlah panggilan yang diterima, jumlah perubahan panggilan, perubahan durasi panggilan yang diterima.

Data seperti detail demografi, tarif pemakaian perorangan tidak digunakan, sehingga hasil yang diperoleh untuk mengelompokkan pelanggan yang berpotensi menjadi tidak loyal dapat dideteksi secara efisien menggunakan *framework* ini sebelum kelompok tersebut masuk ke kelompok yang berpeluang menjadi pelanggan yang loyal. Hasil dari eksperimen terhadap dataset yang ada menunjukkan bahwa pelanggan yang berpotensi untuk *churn* adalah pelanggan yang jumlah pemakaian teleponnya sedikit dan layanan teleponnya juga sedikit seperti sms, mereka inilah yang masuk ke kelompok pelanggan yang tidak loyal (Richter, Yom-Tov, & Slonim, 2010). Loyalitas pelanggan yang akan diukur dalam penelitian ini adalah seberapa banyak pelanggan yang tidak loyal dan benar-benar tidak dapat dipertahankan. Ada beberapa metode untuk mengatasi pelanggan yang tidak loyal, seperti melakukan retensi untuk meningkatkan loyalitas, serta

perusahaan harus memiliki program yang dapat secara akurat membedakan pelanggan yang berisiko tidak loyal dengan pelanggan yang loyal, agar biaya retensi tidak membesar.

2.2.2 SUPPORT VECTOR MACHINE

Support vector machine (SVM) dikembangkan oleh Boser, Guyon, Vapnik dan pertama kali dipresentasikan pada tahun 1992 di *Annual Workshop on Computational Learning Theory*. Konsep SVM dapat dijelaskan secara sederhana sebagai usaha mencari hyperplane terbaik yang berfungsi sebagai pemisah dua buah class pada input space. Gambar 4.1-a memperlihatkan beberapa pattern yang merupakan anggota dari dua buah class : +1 dan -1. Pattern yang tergabung pada class -1 disimbolkan dengan warna merah (kotak), sedangkan pattern pada class +1, disimbolkan dengan warna kuning (lingkaran). Problem klasifikasi dapat diterjemahkan dengan usaha menemukan garis (hyperplane) yang memisahkan antara kedua kelompok tersebut (Cortes & Vapnik, 1995). Garis pemisah (discrimination boundaries) ditunjukkan pada Gambar 4.1-a merupakan salah satu alternatif garis pemisah yang memisahkan kedua class.



Gambar 4.1 Hyperplane Terbaik yang Memisahkan Kedua Class -1 dan +1 (Nugroho, 2008)

Hyperplane pemisah terbaik antara kedua class dapat ditemukan dengan mengukur margin hyperplane tersebut dan mencari titik maksimalnya. Margin adalah jarak antara hyperplane tersebut dengan pattern terdekat dari masing-masing class. Pattern yang paling dekat ini disebut sebagai support vektor. Garis solid pada Gambar 4.1-b menunjukkan hyperplane yang terbaik, yaitu yang terletak tepat pada tengah-tengah kedua class, sedangkan titik merah dan kuning yang berada dalam lingkaran hitam adalah support vector. Usaha untuk mencari lokasi hyperplane ini merupakan inti dari proses pembelajaran pada support vector machine (Nugroho, 2008). Karena dengan menemukan hyperplane terbaik tersebut bisa meningkatkan keakuratan dalam memprediksi loyalitas pelanggan.

Data yang tersedia dinotasikan sebagai  $\vec{X}_i \in R^d$  sedangkan label masing-masing dinotasikan  $y_i \in \{-1 + 1\}$  untuk  $i = 1, 2, \dots, l$ , yang mana  $l$  adalah banyaknya data. Diasumsikan kedua class -1 dan +1 dapat terpisah secara sempurna oleh hyperplane berdimensi  $d$ , yang didefinisikan (Nugroho, 2008):

$$\vec{W} \cdot \vec{X}_i + b = 0 \dots\dots\dots (1)$$

Pattern  $\vec{X}_i$  yang termasuk class -1 (sampel negatif) dapat dirumuskan sebagai pattern yang memenuhi pertidaksamaan

$$\vec{W} \cdot \vec{X}_i + b \leq -1 \dots\dots\dots (2)$$

Sedangkan patter  $\vec{X}_i$  yang termasuk class +1 (sampel positif)

$$\vec{W} \cdot \vec{X}_i + b \geq +1 \dots\dots\dots (3)$$

Margin terbesar dapat ditemukan dengan memaksimalkan nilai jarak antara hyperplane dan titik terdekatnya, yaitu  $1/\|\vec{w}\|$ . Hal ini dapat dirumuskan sebagai Quadratic Programming (QP) problem, yaitu mencari titik minimal persamaan, dengan memperhatikan constraint persamaan.

$$\min_{\vec{w}} \tau(w) = \frac{1}{2} \|\vec{w}\|^2 \dots\dots\dots (4)$$

$$y_i(\vec{x}_i \cdot \vec{w} + b) - 1 \geq 0, \forall i \dots\dots\dots (5)$$

Problem ini dapat dipecahkan dengan berbagai teknik komputasi, diantaranya Lagrange Multiplier

$$L(\vec{w}, b, a) = \frac{1}{2} \|\vec{w}\|^2 - \sum_{i=1}^l a_i (y_i(\vec{x}_i \cdot \vec{w} + b) - 1) \quad (i = 1, 2, \dots, l) \dots\dots\dots (6)$$

$a_i$  adalah Lagrange multipliers, yang bernilai nol atau positif ( $a_i \geq 0$ ). Nilai optimal dari persamaan dapat dihitung meminimalkan  $L$  terhadap  $\vec{w}_i$  dan  $b$ , dan dengan memaksimalkan  $L$  terhadap  $a_i$ . Dengan memperhatikan sifat bahwa pada titik optimal gradient  $L = 0$ , persamaan dapat dimodifikasi sebagai maksimalisasi problem yang hanya mengandung  $a_i$  saja, sebagaimana persamaan  $\alpha$  di bawah.

Maximize:

$$\sum_{i=1}^l a_i - \frac{1}{2} \sum_{i,j=1}^l a_i a_j y_i y_j \vec{x}_i \cdot \vec{x}_j \dots\dots\dots (7)$$

Subject to:

$$a_i \geq 0 (i = 1, 2, \dots, l) \quad \sum_{i=1}^l a_i y_i = 0 \dots\dots\dots (8)$$

Dari hasil perhitungan ini diperoleh  $i$  yang  $a_i$  kebanyakan bernilai positif. Data yang berkorelasi dengan  $a_i$  yang positif inilah yang disebut sebagai support vector (Cortes & Vapnik, 1995).

2.2.3 KERNEL TRIKS

Pada umumnya permasalahan dalam domain dunia nyata (real world problem) datanya sangat sedikit sekali yang bersifat linear, kebanyakan data bersifat non-linear. Untuk menyelesaikan kasus non-linear, perhitungan SVM dimodifikasi menjadi dua tahap, dimana didalamnya memanfaatkan konsep yang disebut Kernel trick (Nugroho, 2008). Ide yang mendasarinya adalah mengubah data bersifat non-linear dan dimensi dari feature space cukup tinggi, maka data pada input space dapat dipetakan ke feature space yang baru, dimana pola-pola tersebut pada probabilitas tinggi dapat dipisahkan secara linear. Untuk menyelesaikan masalah non-linear, support vector machine dimodifikasi dengan memasukkan kernel trick yang mengubah data non-linear ke data linear (Hamel, 2009). Kernel trick dapat dirumuskan dengan:

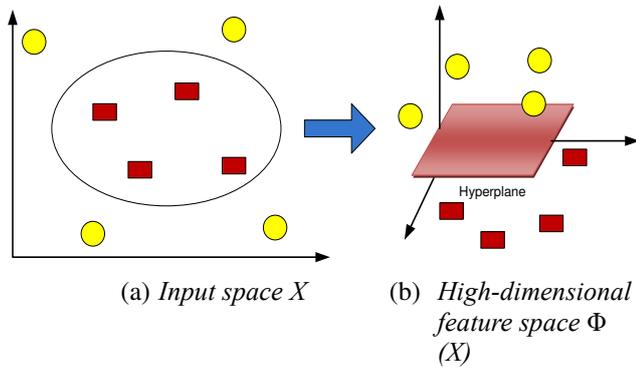
$$K(\vec{X}_i, \vec{X}_j) = \Phi(\vec{X}_i) \cdot \Phi(\vec{X}_j) \dots\dots\dots (9)$$

Ilustrasi dari konsep ini dapat dilihat pada gambar 2.6. Pada Gambar 2.6-a diperlihatkan data pada class kuning dan data pada class merah yang berada pada input space berdimensi dua tidak dapat dipisahkan secara linear. Selanjutnya Gambar 2.6-b menunjukkan bahwa fungsi  $\Phi$  memetakan tiap data pada

*input space* tersebut ke ruang vector baru yang berdimensi lebih tinggi (dimensi 3), dimana kedua *class* dapat dipisahkan secara *linear* oleh sebuah *hyperplane*. Notasi matematika dari *mapping* ini adalah sebagai berikut:

$$\Phi : R^d \rightarrow R^q \quad d < q \dots\dots\dots (10)$$

Pemetaan ini dilakukan dengan menjaga topologi data, dalam artian dua data yang berjarak dekat pada *input space* akan berjarak dekat juga pada *feature space*, sebaliknya dua data yang berjarak jauh pada *input space* akan juga berjarak jauh pada *feature space*.



Gambar 2.1 Kedua *Class* Dipisahkan Secara *Linear* oleh Sebuah *Hyperplane* (Nugroho, 2008)

Selanjutnya proses pembelajaran pada SVM dalam menemukan titik-titik *support vector*, hanya bergantung pada *dot product* dari data yang sudah berdimensi lebih tinggi, yaitu  $\Phi(\vec{X}_i) \cdot \Phi(\vec{X}_j)$ . Karena umumnya transformasi  $\Phi$  ini tidak diketahui, dan sangat sulit untuk dipahami secara mudah, maka perhitungan *dot product* tersebut dapat digantikan dengan fungsi kernel  $K(\vec{X}_i, \vec{X}_j)$  yang mendefinisikan secara implisit transformasi  $\Phi$ . *Kernel trick* memberikan berbagai kemudahan, karena dalam proses pembelajaran SVM, untuk menentukan *support vector*, hanya cukup mengetahui fungsi kernel yang dipakai, dan tidak perlu mengetahui wujud dari fungsi *non-linear*  $\Phi$ . Selanjutnya hasil klasifikasi dari data  $\vec{X}$  diperoleh dari persamaan berikut (Cortes & Vapnik, 1995):

$$f(\Phi(\vec{X})) = \vec{w} \cdot \Phi(\vec{X}) + b \dots\dots\dots (11)$$

$$= \sum_{i=1, \vec{X}_i \in SV}^n a_i y_i \Phi(\vec{X}) \cdot \Phi(\vec{X}_i) + b \dots\dots\dots (12)$$

$$= \sum_{i=1, \vec{X}_i \in SV}^n a_i y_i K(\vec{X}, \vec{X}_i) + b \dots\dots\dots (13)$$

Support vektor pada persamaan di atas dimaksudkan dengan subset dari training set yang terpilih sebagai support vector, dengan kata lain data  $\vec{X}_i$  yang berkorespondensi pada  $a_i \geq 0$ . Fungsi kernel yang biasanya dipakai dalam literatur SVM yaitu (Nugroho, 2008):

- a. Linear :  $K(x, y) = x \cdot y$
- b. Polynomial :  $K(x, y) = (x \cdot y + 1)^d$
- c. Radial Basis Funtion (RBF):  

$$K(x, y) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right)$$
- d. Sigmoid :  $K(x, y) = \tanh(\beta x^T x_i + \beta_1)$ , dimana  $\beta, \beta_1 \in \mathcal{R}$

2.2.4 BOOTSTRAPPING

Bootstrap adalah metode *resampling* yang telah diterapkan secara luas dan memungkinkan pembuatan model yang lebih realistis (Efron & Tibshirani, 1998). Sebelumnya setiap kali

sampel yang diambil dari dataset untuk membentuk data *training* atau *testing* itu diambil tanpa penggantian, artinya contoh yang sama setelah dipilih tidak dapat dipilih lagi (Witten, I. H., Frank, E., & Hall, 2011), jadi dengan menggunakan bootstrapping sampel yang sudah dipilih dapat dipilih kembali, hal ini memungkinkan penilaian terhadap estimasi akurasi dengan random sampling dengan penggantian dari dataset yang asli sehingga bisa meningkatkan akurasi dan mempercepat waktu komputasinya.

Metode bootstrap dilakukan dengan mengambil sampel dari sampel yang asli dengan ukuran sama dengan aslinya dan dilakukan dengan penggantian, biasanya ukuran resampling diambil secara ribuan kali agar dapat mewakili data populasinya. Kedudukan sampel asli dalam metode bootstrap dipandang sebagai populasi. Metode penyampelan ini biasa disebut dengan *resampling bootstrap with replacement* (Tian, Song, Li, & de Wilde, 2014). Bootstrap dapat digunakan untuk mengatasi permasalahan statistik baik dalam masalah data yang sedikit, data yang menyimpang dari asumsinya maupun data yang tidak memiliki asumsi dalam distribusinya. Dalam beberapa aplikasi yang mengolah data yang besar biasanya menggunakan metode sampling (Witten, I. H., Frank, E., & Hall, 2011) untuk menyaring agar data yang diproses dapat menjadi lebih kecil.

Metode distribusi populasi didasarkan pada prinsip bootstrap yang dapat digunakan prinsip plug-in untuk memperkirakan parameter dari sampel. Untuk memperkenalkan secara singkat prinsip plug-in yang diasumsikan bahwa  $\theta$  adalah parameter dan F adalah kemungkinan distribusi. Tujuannya adalah untuk menghitung  $\theta$  dengan menerapkan beberapa prosedur numerik  $t(\cdot)$  yaitu:

$$\theta = t(F) \dots\dots\dots (14)$$

Oleh karena itu estimasinya dapat didefinisikan sebagai

$$\hat{\theta} = t(\hat{F}) \dots\dots\dots (15)$$

dimana  $\hat{F}$ ) fungsi distribusi empirik dari sampel random (Tian et al., 2014). Dalam hal ini parameter  $\theta$  akan menjadi indeks sensitivitas atau kesalahan prediksi dalam menilai pelanggan yang loyal atau tidak loyal. Secara ringkas langkah-langkah bootstrap adalah sebagai berikut (Efron & Tibshirani, 1998):

1. Menentukan jumlah  $B$  sampel independen bootstrap  $X^*1, X^*2, \dots, X^*B$  di mana masing-masing sampel berisi  $n$  data yang diperoleh dari  $x$  (data awal).
2. Mengevaluasi replikasi yang ada pada masing-masing sampel bootstrap.
3. Mengestimasi sampel dengan menggunakan standar deviasi untuk bootstrap yang direplikasi  $B$  kali.

2.2.5 SELEKSI FITUR FORWARD SELECTION

Setelah dilakukan *sampling* pada dataset, kemudian data difilter kembali dengan menggunakan metode seleksi fitur forward selection. Seleksi fitur adalah salah satu teknik terpenting dan sering digunakan dalam *pre-processing data mining* khususnya untuk *knowledge discovery*. Teknik ini mengurangi atau mereduksi jumlah fitur yang terlibat dalam menentukan suatu nilai kelas target, mengurangi fitur yang tidak relevan (Larose, 2007) dan data yang menyebabkan salah pengertian terhadap kelas target yang membuat efek bagi aplikasi.

2.2.6 WEIGHTED INFORMATION GAIN

Weighted Information Gain (WIG) sering digunakan untuk meranking atribut yang paling berpengaruh terhadap kelasnya. Nilai gain dari suatu atribut, diperoleh dari nilai entropi sebelum pemisahan dikurangi dengan nilai entropi setelah pemisahan. Tujuan pengurangan fitur pengukuran nilai informasi diterapkan sebagai tahap sebelum pengolahan awal. Hanya atribut memenuhi kriteria (*threshold*) yang ditentukan dipertahankan untuk digunakan oleh algoritma klasifikasi (Bramer, 2007). 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 (Gallager, 2001) dituliskan sebagai (Han, J., & Kamber, 2012):

$$Info(D) = - \sum_{i=1}^m p_i \log_2(p_i) \dots\dots\dots (16)$$

Berdasarkan persamaan di atas berikut merupakan keterangannya:

- D : Himpunan Kasus
- m : Jumlah partisi D
- $p_i$  : Proporsi dari  $D_i$  terhadap D

Dalam hal ini  $p_i$  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 nilai entropi setelah pemisahan dengan cara sebagai berikut (Han, J., & Kamber, 2012):

$$Info_A(D) = - \sum_{j=1}^v \frac{|D_j|}{|D|} \times Info(D_j) \dots\dots\dots (17)$$

Berdasarkan persamaan di atas berikut merupakan keterangannya:

- D : Himpunan kasus
- A : Atribut
- v : Jumlah partisi atribut A
- $|D_j|$  : Jumlah kasus pada partisi ke j
- $|D|$  : Jumlah kasus dalam D
- $Info(D_j)$  : Total entropi dalam partisi

Untuk mencari nilai information gain atribut A diperoleh dengan persamaan berikut(Han, J., & Kamber, 2012):

$$Gain(A) = Info(D) - Info_A(D) \dots\dots\dots (18)$$

Berdasarkan persamaan di atas berikut merupakan keterangannya:

- Gain(A) : Information atribut A
- Info(D) : Total entropi
- $Info_A(D)$  : Entropi A

Dengan penjelasan lain, Gain(A) adalah reduksi yang diharapkan didalam 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

3 METODE PENELITIAN

3.1 DATASET

Dalam penelitian ini dikumpulkan dua dataset, dataset yang pertama dataset *churn* dari database *University of California* (UCI) dan dataset yang kedua dataset *telecom* dari *Customer Relationship Management (CRM) of Duke University* seperti pada Tabel 3.1. Dataset *churn* dapat diunduh disitus <http://www.sgi.com/tech/mlc/db/> dan dataset *telecom* dapat diunduh melalui situs <http://www.fuqua.duke.edu/centers/ccrm/datasets/download.html>. Kedua dataset ini banyak digunakan oleh para peneliti sebelumnya (Xia & Jin, 2008) (Z.-Y. Chen et al., 2012) karena dataset *churn* dan *telecom* ini merupakan dataset publik dibidang telekomunikasi yang banyak digunakan oleh para peneliti dibidang loyalitas pelanggan.

Tabel 3.1 Dataset Loyalitas Pelanggan

No	Name	Type	Record	Dimension
1	Churn	Churn	4.974	21
2	Telecom	Churn	3.399	6

Pada dataset *telecom* karena jumlah *record* aslinya terlalu besar yaitu sebanyak 195.956, maka dataset *telecom* hanya akan diambil *record* sebanyak 3.399. Jumlah ini sudah melibihi batas minimal untuk porsi yang ideal sebagaimana yang telah diterapkan dalam ilmu statistik, dimana pengambilan sebagian dataset dapat mewakili jumlah dari populasi aslinya dapat dirumuskan dengan formula sebagai berikut (Kriyantono, 2008):

$$n = \frac{N}{N.d^2+1} \dots\dots\dots (31)$$

Di mana:

- n : Ukuran sampel
- N : Ukuran populasi
- d : Persen kelonggaran ketidaktelitian karena kesalahan pengambilan sampel yang masih dapat ditolerir.

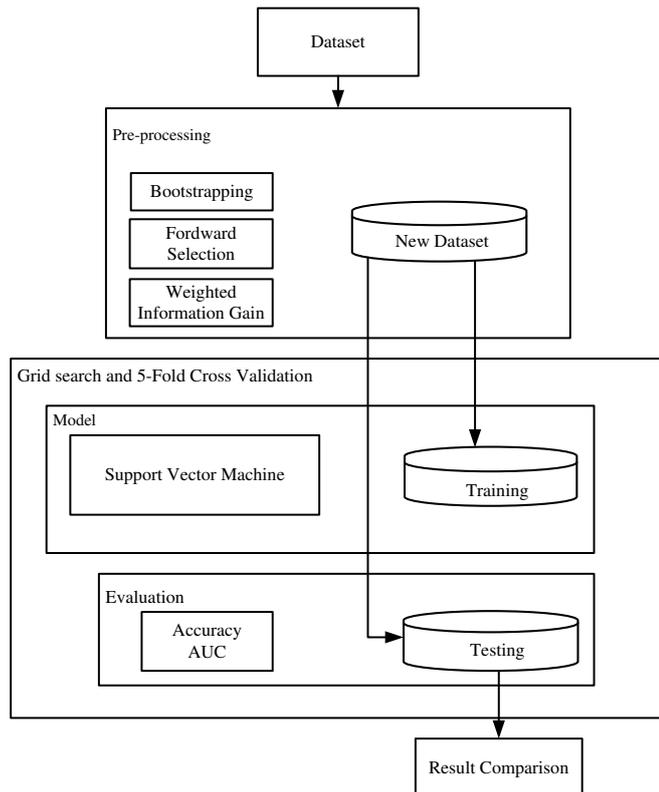
3.2 METODE YANG DIUSULKAN

Untuk mengatasi masalah ketidakseimbangan kelas, memfilter atribut yang tidak relevan serta kelemahan algoritma SVM yaitu sulitnya menentukan nilai parameter dan pemilihan fungsi kernel yang tepat, dalam penelitian ini diusulkan dengan menggunakan metode bootstrapping, FS-WIG dan *grid search* pada algoritma SVM dengan fungsi kernel Radial Basis Function (RBF). Metode yang diusulkan sebagaimana yang ditunjukkan pada Gambar 3.1 yaitu:

1. Pada tahap pertama yaitu dengan *pre-processing* dengan metode sampel bootstrapping.
2. Kemudian data hasil sampel bootstrapping difilter kembali menggunakan seleksi fitur Forward Selection (FS) dan kemudian atribut atau fitur diberi pembobotan dengan menggunakan metode *Weighted Information Gain* (WIG). Kemudian fitur-fitur tersebut dihitung nilai gainnya, setelah semua fitur diketahui nilai gainnya langkah selanjutnya adalah menentukan nilai ambang yaitu nilai batas antara fitur terpilih dengan fitur yang tidak terpilih.
3. Kemudian metode *grid search* digunakan untuk memilih nilai parameter yang paling optimal yaitu dengan memasukkan nilai parameter  $C = 0,001$  dan  $1,5$ ,  $\epsilon = 0.004 - 1.5$  dan  $\gamma = 0.001 - 1.5$ .
4. Setelah itu dataset dibagi dengan metode *5 fold-cross validation* yaitu data *training* dan data *testing*.

5. Kemudian diklasifikasi dengan algoritma support vector machine (LibSVM) sebagai usaha untuk mencari *support vector* sebagai batas *margin* yang memisahkan kedua kelas.
6. Yang terakhir adalah tahapan pengukuran akurasi prediksi dengan melihat performa akurasi dan AUC.

Dalam penelitian pada umumnya pengujian nilai k pada *cross validation* dilakukan sebanyak 10 kali untuk memperkirakan akurasi estimasi, tetapi dalam penelitian ini nilai k yang digunakan berjumlah 5 atau *5-fold cross validation*, hal ini sengaja dilakukan karena pengujian dengan *5-fold cross validation* dapat mempercepat waktu komputasinya mengingat dataset yang digunakan cukup besar.



Gambar 3.1 Diagram Metode yang Diusulkan

## 4 HASIL EKSPERIMEN

### 4.1 HASIL EKSPERIMEN DATASET CHURN

Dalam melakukan penelitian ini digunakan spesifikasi komputer dengan processor Intel® Core™ i3-2330M, 2.20GHz, memory 6 GB, hardisk 500 GB, sistem operasi Windows 7 Ultimate SP-1 64-bit dan aplikasi *RapidMiner* 5.3.015.

#### 4.1.1 HASIL EKSPERIMEN BOOTSTRAPPING DENGAN SVM

Hasil eksperimen terbaik sampel bootstrapping dengan SVM yang telah dilakukan dengan menginputkan nilai parameter C = 0.5, gamma 0.001 dan epsilon = 0.004 yang menghasilkan nilai akurasi = 87.14% sebagaimana ditunjukkan dalam Tabel 4.1.

Tabel 4.1 Komparasi Algoritma SVM dan SVM+Bootstrapping pada Dataset Churn

SVM	85.87%	0.504
<b>SVM dengan Bootstrapping</b>	<b>87.14%</b>	<b>0.500</b>

#### 4.1.2 HASIL EKSPERIMEN BOOTSTRAPPING, FS-WIG DENGAN SVM

Eksperimen selanjutnya dengan menggunakan bootstrapping, FS-WIG pada SVM, dengan menginputkan parameter SVM dengan nilai parameter C = 0,5, gamma 0.001 dan epsilon = 0.004. Hasil eksperimen terbaik sebagaimana yang ditunjukkan oleh Tabel 4.2.

Tabel 4.2 Komparasi Algoritma SVM dan SVM+Bootstrapping+FS-WIG pada Dataset Churn

	Akurasi	AUC
SVM	85.87%	0.504
<b>SVM dengan Bootstrapping dan FS-WIG</b>	<b>91.52%</b>	<b>0.762</b>

#### 4.1.3 HASIL EKSPERIMEN BOOTSTRAPPING, FS-WIG DAN GRID SEARCH DENGAN SVM

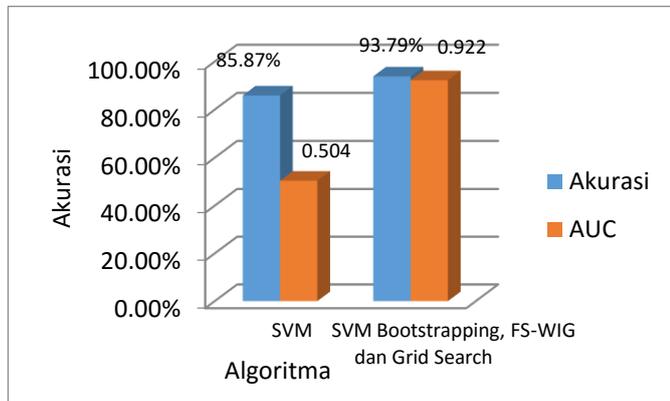
Pada pengujian dataset *churn* ini dilakukan eksperimen dengan memakai kernel Radial Basis Function (RBF), kemudian memasukkan nilai parameter C, epsilon dan gamma. Algoritma bootstrapping menggunakan sampel *relative* dan parameter sampel *ratio* 0.1, parameter ini berfungsi menginputkan dari sebagian kecil data dari jumlah total dataset yang digunakan dan memberikan bobot secara manual pada sampel *ratio*. Hasil eksperimen terbaik yang telah dilakukan untuk penentuan nilai akurasi dan AUC adalah dengan parameter C secara *logarithmic* antara 0.001 – 1.5, gamma = 0.001 – 1.5 dan epsilon = 0.004 – 1.5 yang menghasilkan nilai akurasi = 93.79% dan AUC nya adalah 0.922.

Pada Tabel 4.3 dari eksperimen yang telah dilakukan dan rata-rata keseluruhan eksperimen pada dataset *churn* secara konsisten menunjukkan peningkatan nilai akurasi dan AUC yang signifikan antara yang menggunakan SVM dengan sampel bootstrapping, FS-WIG dan *grid search* pada algoritma SVM. Gambar 4.1 merupakan grafik komparasi antara algoritma SVM dengan bootstrapping, FS-WIG dan *grid search* pada SVM

Tabel 4.3 Komparasi Algoritma SVM dengan Bootstrapping, FS-WIG dan Grid Search pada Dataset Churn

	Akurasi	AUC
SVM	85.87%	0.504
<b>SVM dengan Bootstrapping, FS-WIG serta Grid Search.</b>	<b>93.79%</b>	<b>0.922</b>

	Akurasi	AUC
--	---------	-----



Gambar 4.1 Grafik Akurasi dan AUC Algoritma SVM dan SVM dengan Bootstrapping, WIG serta *Grid Search* pada Dataset *Churn*

## 4.2 HASIL EKSPERIMEN DATASET TELECOM

### 4.2.1 HASIL EKSPERIMEN BOOTSTRAPPING DENGAN SVM

Hasil eksperimen, setelah dataset baru dipilih sampel bootstrapping kemudian dengan menginputkan nilai parameter SVM, yaitu parameter C = 0.05, gamma = 0.001 dan epsilon = 0.004 dan kemudian diambil nilai rata-ratanya. Hasil eksperimen sebagaimana ditunjukkan oleh Tabel 4.4.

Tabel 4.4 Komparasi Algoritma SVM dan SVM dengan Bootstrapping pada Dataset *Telecom*

	Akurasi	AUC
SVM	97.03%	0.610
<b>SVM dengan Bootstrapping</b>	<b>98.56%</b>	<b>0.870</b>

### 4.2.2 HASIL EKSPERIMEN BOOTSTRAPPING, FS-WIG DENGAN SVM

Hasil eksperimen selanjutnya, setelah dataset baru dipilih sampel bootstrapping kemudian atribut diseleksi fitur dengan menggunakan FS-WIG, kemudian menginputkan nilai parameter SVM, yaitu parameter C = 0.05, gamma = 0.001 dan epsilon = 0.004 dan kemudian diambil nilai rata-ratanya. Hasil eksperimen sebagaimana ditunjukkan pada Tabel 4.5.

Tabel 4.5 Komparasi Algoritma SVM dan SVM dengan Bootstrapping dan FS-WIG pada Dataset *Telecom*

	Akurasi	AUC
SVM	97.03%	0.610
<b>SVM dengan Bootstrapping dan FS-WIG</b>	<b>98.56%</b>	<b>0.870</b>

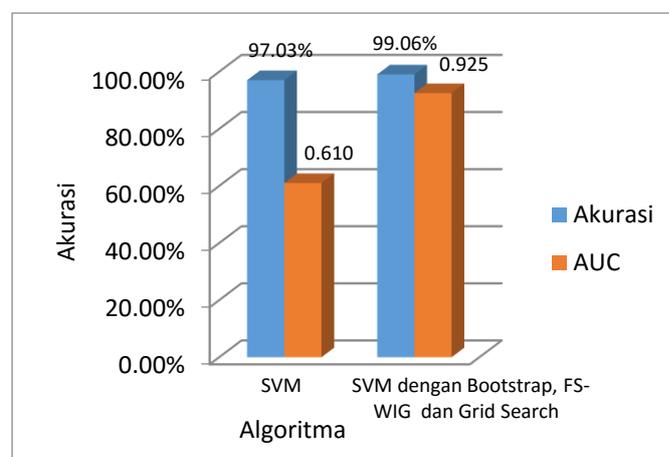
### 4.2.3 HASIL EKSPERIMEN BOOTSTRAPPING, FS-WIG DAN GRID SEARCH DENGAN SVM

Sedangkan pengujian pada dataset *telecom* dilakukan dengan memakai kernel Radial Basis Function (RBF), kemudian memasukkan nilai parameter C dan epsilon. Pada pengujian ini dilakukan eksperimen dengan memakai kernel Radial Basis Function (RBF), kemudian memasukkan nilai parameter C, epsilon dan gamma. Algoritma bootstrapping menggunakan sampel *relative* dan parameter sampel *ratio* 0.1, parameter ini berfungsi menginputkan dari sebagian kecil data

dari jumlah total dataset yang digunakan dan memberikan bobot secara manual pada sampel *ratio*. Hasil eksperimen terbaik yang telah dilakukan untuk penentuan nilai akurasi dan AUC adalah dengan parameter C secara *logarithmic* antara 0,001 – 1,5, gamma = 0,001 – 1,5 dan epsilon = 0,004 – 1,5 yang menghasilkan nilai akurasi = 99.96% dan AUC nya adalah 0.925%, sebagaimana yang telah ditunjukkan pada Tabel 4.6. Gambar 4.2 merupakan grafik komparasi antara algoritma SVM dengan bootstrapping, FS-WIG dan *grid search* pada SVM

Tabel 4.6 Komparasi Algoritma SVM dan SVM dengan Bootstrapping, FS-WIG serta *Grid Search* pada Dataset *Telecom*

	Akurasi	AUC
SVM	97.03%	0.610
<b>SVM dengan Bootstrapping, FS-WIG serta <i>Grid Search</i></b>	<b>99.06%</b>	<b>0.925</b>



Gambar 4.2 Grafik akurasi dan AUC Algoritma SVM dan SVM dengan Bootstrapping, FS-WIG serta *Grid Search* pada Dataset *Telecom*

## 4.3 HASIL KOMPARASI DENGAN PENELITIAN TERDAHULU

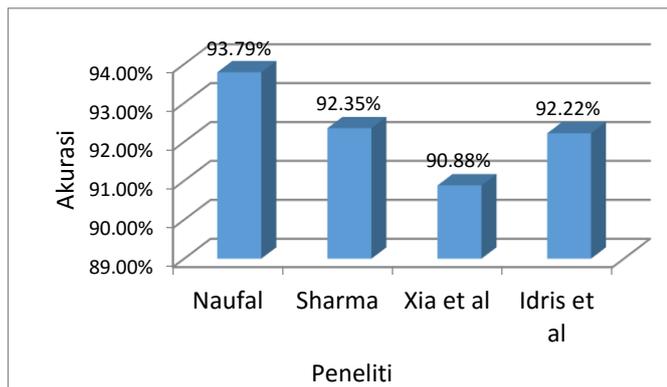
Dari eksperimen yang telah peneliti lakukan membuktikan bahwa model yang peneliti usulkan ini lebih unggul dari pada model yang telah dilakukan oleh Xia et al (Xia & Jin, 2008), yang mana dataset *churn* yang peneliti gunakan sama dengan dataset yang digunakan oleh Xia *et al* dengan rata-rata akurasi 90.88%. Penelitian selanjutnya juga dilakukan oleh Sharma (Sharma & Panigrahi, 2011) pada tahun 2011 yang juga menggunakan dataset *churn* dengan menerapkan model *Feed-Forward Back-Propagation Neural Network*, untuk mencegah masalah *over training* pada neural network, Sharma mengambil data *training* secara acak tetapi setelah melewati beberapa kali jaringan, metode yang diusulkan Sharma malah kehilangan kemampuan generalisasi sehingga tetap terjebak dalam permasalahan *over training* tersebut, dan metode Sharma dalam memprediksi loyalitas pelanggan ini memperoleh rata-rata akurasi 92.35%, sementara dalam penelitian ini dataset *churn* memperoleh rata-rata akurasi 93.79%.

Metode yang diusulkan dalam eksperimen ini juga lebih baik dibanding dengan metode yang diusulkan oleh Idris et al (Idris et al., 2012) yang menggunakan *undersampling* Particle Swarm Optimization (PSO) dengan algoritma random forest dan fitur seleksi Principle Component Analysis (PCA) dan

menggunakan dataset *french telecom company* yang hasilnya memperoleh tingkat rata-rata akurasi 92.22%, dengan demikian model yang diusulkan dalam penelitian ini lebih baik dengan memperoleh rata-rata akurasi 93.79% untuk dataset *churn* dan 99.06% untuk dataset *telecom*. Tabel 4.13 menunjukkan komparasi model yang telah diusulkan dalam penelitian ini dengan beberapa model peneliti terdahulu.

Tabel 4.7 Komparasi Performansi dengan Model yang Diusulkan Peneliti Terdahulu

Peneliti	Model	Dataset	Hasil
Naufal	Sampel Bootstrapp, FS-WIG dan <i>grid search</i> pada SVM dengan RBF Kernel	<i>Churn</i>	93.79%
		<i>Telecom</i>	99.06%
Sharma	<i>Feed-Forward Back-Propagation Neural Network</i>	<i>Churn</i>	92.35%
Xia et al	SVM dengan RBF Kernel	<i>Churn</i>	90.88%
Idris et al	Undersampling PSO, Random Forest dan fitur seleksi PCA	<i>French Telecom Company</i>	92.22%



Gambar 4.1 Grafik Komparasi Performansi dengan Model yang Diusulkan Peneliti Terdahulu

## 5 KESIMPULAN

Beberapa eksperimen dilakukan untuk mendapatkan arsitektur yang optimal dan menghasilkan prediksi yang akurat. Berdasarkan hasil eksperimen, dapat disimpulkan sebagai berikut:

1. Hasil eksperimen pada dataset *churn* dengan SVM standar didapatkan nilai akurasi 85.87% dan AUC 0.504, sedangkan dengan menggunakan metode SVM+bootstrapping dengan nilai parameter  $C = 0,05$ ,  $\epsilon = 0,004$  dan  $\gamma = 0,001$  didapatkan nilai akurasi 87.14% dan AUC 0.500. Sedangkan hasil eksperimen SVM standar pada dataset *telecom* didapatkan akurasi 97.03% dan AUC 0.610, sedangkan dengan SVM+bootstrapping dengan parameter  $C = 0,05$ ,  $\epsilon = 0,004$  dan  $\gamma = 0,001$  didapatkan akurasi 98.56% dan AUC 0.870. Dengan menggunakan SVM+bootstrapping ada kenaikan akurasi dan AUC yang konsisten jika diterapkan pada dataset *churn* maupun dataset *telecom*, dimana dataset *churn* dan *telecom*

merupakan dataset yang mengandung ketidakseimbangan kelas. Jadi dapat disimpulkan bahwa metode SVM dengan bootstrapping dalam eksperimen ini bisa untuk mengatasi permasalahan ketidakseimbangan kelas sehingga bisa meningkatkan performansi SVM.

2. Hasil eksperimen selanjutnya pada dataset *churn* dengan SVM standar didapatkan nilai akurasi 85.87% dan AUC 0.504, sedangkan dengan menggunakan metode SVM+bootstrapping+FS-WIG dengan nilai parameter  $C = 0,05$ ,  $\epsilon = 0,004$  dan  $\gamma = 0,001$  didapatkan nilai akurasi 91.52% dan AUC 0.762. Sedangkan hasil eksperimen SVM standar pada dataset *telecom* didapatkan akurasi 97.03% dan AUC 0.610, sedangkan dengan SVM+bootstrapping+FS-WIG dengan parameter  $C = 0,05$ ,  $\epsilon = 0,004$  dan  $\gamma = 0,001$  didapatkan akurasi 98.56% dan AUC 0.870. Dengan demikian SVM+bootstrapping+FS-WIG ada kenaikan akurasi dan AUC yang konsisten jika diterapkan pada dataset *churn*, tetapi pada dataset *telecom* tidak menunjukkan ada peningkatan performansi dari eksperimen SVM+bootstrapping. Hal ini dikarenakan atribut dalam dataset *telecom* hanya 6 atribut sehingga algoritma FS-WIG tidak terlalu berpengaruh jika diterapkan pada dataset yang mempunyai atribut sedikit. Sedangkan pada dataset *churn* ada peningkatan akurasi dan AUC, hal ini dikarenakan dataset *churn* mempunyai atribut yang lebih banyak daripada dataset *telecom*, yaitu 21 atribut. Dari sini dapat disimpulkan bahwa metode fitur seleksi FS-WIG tidak cocok jika diterapkan pada dataset yang mempunyai atribut sedikit, tetapi sangat baik jika diterapkan pada dataset yang berdimensi tinggi dimana dataset yang berdimensi tinggi secara umum mengandung atribut yang kurang relevan.
3. Hasil pengujian pada dataset *churn* dengan SVM standar didapatkan nilai akurasi 85.87% dan AUC 0.504, sedangkan dengan menggunakan metode SVM dengan bootstrapping, FS-WIG serta *grid search* dengan nilai parameter secara *logarithmic*  $C = 0,001 - 1,5$ ,  $\epsilon = 0,004 - 1,5$  dan  $\gamma = 0,001 - 1,5$  didapatkan nilai akurasi 93.79% dan AUC 0.922. Sedangkan hasil pengujian SVM standar pada dataset *telecom* didapatkan akurasi 97.03% dan AUC 0.610, sedangkan SVM dengan bootstrapping, FS-WIG dan *grid search* dengan parameter secara *logarithmic*  $C = 0,001 - 1,5$ ,  $\epsilon = 0,004 - 1,5$  dan  $\gamma = 0,001 - 1,5$  didapatkan akurasi 99.06% dan AUC 0.925. Dengan menggunakan SVM dengan sampel bootstrapping, FS-WIG dan *grid search* ada kenaikan akurasi dan AUC yang konsisten jika diterapkan pada dataset *churn* maupun dataset *telecom*. Hal ini dikarenakan pemilihan parameter SVM secara *grid search* dapat mencari nilai parameter terbaik yang nilainya antara 0,001 - 1,5 untuk parameter C, parameter  $\epsilon$  0,004 - 1,5 dan parameter  $\gamma = 0,001 - 1,5$ . Jadi dapat disimpulkan bahwa metode *grid search* untuk pemilihan parameter SVM dalam eksperimen ini bisa mengatasi permasalahan sulitnya pemilihan nilai parameter pada SVM sehingga bisa memudahkan dalam menentukan nilai parameter yang tepat untuk algoritma SVM.

Dari hasil pengujian ini maka secara umum dapat dianalisa bahwa ada kenaikan akurasi sebesar 7.92% dan AUC sebesar 0.418 pada dataset *churn*, dan kenaikan akurasi sebesar 2.03% dan AUC sebesar 0.315 pada dataset *telecom* setelah diterapkan metode yang diusulkan. Dengan demikian dapat disimpulkan bahwa metode sampel bootstrapping dapat

mengatasi permasalahan pada ketidakseimbangan kelas (*class imbalance*) yang membantu menaikkan kinerja algoritma SVM dan seleksi fitur FS – WIG mampu memilih atribut terbaik. Sedangkan untuk menentukan parameter dengan metode *grid search* mampu memudahkan dalam pemilihan nilai parameter terbaik pada Algoritma SVM, sehingga menghasilkan kinerja atau tingkat akurasi prediksi dalam loyalitas pelanggan yang lebih baik dibanding dengan menggunakan metode individual algoritma SVM.

## DAFTAR PUSTAKA

- Bergstra, J., & Bengio, Y. (2012). Random Search for Hyper-Parameter Optimization. *Journal of Machine Learning Research*, 13, 281–305.
- Bramer, M. (2007). *Principles of Data Mining*. Springer. Retrieved from <http://link.springer.com/article/10.2165/00002018-200730070-00010>
- Burez, J., & Van den Poel, D. (2009). Handling class imbalance in customer churn prediction. *Expert Systems with Applications*, 36(3), 4626–4636. doi:10.1016/j.eswa.2008.05.027
- Charu C. Aggarwal, P. S. Y. (2008). *Privacy Preserving Data Mining* (Vol. 19). Springer US. doi:10.1007/978-0-387-29489-6
- Chen, H., Zhang, J., Xu, Y., Chen, B., & Zhang, K. (2012). Performance comparison of artificial neural network and logistic regression model for differentiating lung nodules on CT scans. *Expert Systems with Applications*, 39(13), 11503–11509. doi:10.1016/j.eswa.2012.04.001
- Chen, Z., & Fan, Z. (2013). Knowledge-Based Systems Dynamic customer lifetime value prediction using longitudinal data : An improved multiple kernel SVR approach. *Knowledge-Based Systems*, 43, 123–134. doi:10.1016/j.knsys.2013.01.022
- Chen, Z.-Y., Fan, Z.-P., & Sun, M. (2012). A hierarchical multiple kernel support vector machine for customer churn prediction using longitudinal behavioral data. *European Journal of Operational Research*, 223(2), 461–472. doi:10.1016/j.ejor.2012.06.040
- Cortes, C., & Vapnik, V. (1995). Support vector machine. In *Machine learning* (pp. 1303–1308). doi:10.1007/978-0-387-73003-5\_299
- Coussement, K., & Van den Poel, D. (2008). Churn prediction in subscription services: An application of support vector machines while comparing two parameter-selection techniques. *Expert Systems with Applications*, 34(1), 313–327. doi:10.1016/j.eswa.2006.09.038
- Efron, B., & Tibshirani, R. (1998). *An introduction to the bootstrap*. New York: Chapman & Hall Book. Retrieved from <http://books.google.com/books?hl=en&lr=&id=gLlPUIUxRntoC&oi=fnd&pg=PR14&dq=An+Introduction+to+the+Bootstrap&ots=A8wrX6QbF7&sig=6gK8Gx-KtVcUXJM7qSFv92zi3eM>
- Farvaresh, H., & Sepehri, M. M. (2011). A data mining framework for detecting subscription fraud in telecommunication. *Engineering Applications of Artificial Intelligence*, 24(1), 182–194. doi:10.1016/j.engappai.2010.05.009
- Gallager, R. (2001). Claude E. Shannon: A retrospective on his life, work, and impact. *Information Theory, IEEE Transactions on*, 47(7), 2681–2695. Retrieved from [http://ieeexplore.ieee.org/xpls/abs\\_all.jsp?arnumber=959253](http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=959253)
- Hamel, L. (2009). *Knowledge discovery with support vector machines*. John Wiley & Sons, Inc. Retrieved from <http://books.google.com/books?hl=en&lr=&id=WaUnU4pEVVUC&oi=fnd&pg=PT10&dq=Knowledge+Discovery+with+Support+Vector+Machine&ots=U9cp-ZSxZ3&sig=XN99rPTt36-mZO-PpHdhwbhJ9-I>
- Han, S. H., Lu, S. X., & Leung, S. C. H. (2012). Segmentation of telecom customers based on customer value by decision tree model. *Expert Systems with Applications*, 39(4), 3964–3973. doi:10.1016/j.eswa.2011.09.034
- Han, J., & Kamber, M. (2012). *Data Mining: Concepts and Techniques* (3rd Editio.). Morgan Kaufmann Publishers.
- Huang, B., Kechadi, M. T., & Buckley, B. (2012). Customer churn prediction in telecommunications. *Expert Systems with Applications*, 39(1), 1414–1425. doi:10.1016/j.eswa.2011.08.024
- Idris, A., Rizwan, M., & Khan, A. (2012). Churn prediction in telecom using Random Forest and PSO based data balancing in combination with various feature selection strategies. *Computers & Electrical Engineering*, 38(6), 1808–1819. doi:10.1016/j.compeleceng.2012.09.001
- Jadhav, R., & Pawar, U. (2011). Churn Prediction in Telecommunication Using Data Mining Technology. *International Journal of Advanced Computer Science and Applications*, 2(2), 17–19. Retrieved from <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.190.5029&rep=rep1&type=pdf#page=30>
- Khoshgoftaar, T. M., & Gao, K. G. K. (2009). Feature Selection with Imbalanced Data for Software Defect Prediction. *2009 International Conference on Machine Learning and Applications*. doi:10.1109/ICMLA.2009.18
- Kriyantono, R. (2008). *Teknik Praktis Riset Komunikasi*. Jakarta: Kencana. Retrieved from <http://scholar.google.com/scholar?hl=en&btnG=Search&q=in+title:Teknik+Praktis+Riset+Komunikasi#0>
- Larose, D. T. (2007). *Data Mining Methods and Models*. Canada: John Wiley & Sons, Inc.
- Lin, L., Ravitz, G., Shyu, M. L., & Chen, S. C. (2008). Effective feature space reduction with imbalanced data for semantic concept detection. *Proceedings - IEEE International Conference on Sensor Networks, Ubiquitous, and Trustworthy Computing*, 262–269. doi:10.1109/SUTC.2008.66
- Maldonado, S., Weber, R., & Famili, F. (2014). Feature selection for high-dimensional class-imbalanced data sets using Support Vector Machines. *Information Sciences*, 286, 228–246. doi:10.1016/j.ins.2014.07.015
- Mozer, M. C., Wolniewicz, R., Grimes, D. B., Johnson, E., & Kaushansky, H. (2000). Predicting subscriber dissatisfaction and improving retention in the wireless telecommunications industry. *IEEE Transactions on Neural Networks / a Publication of the IEEE Neural Networks Council*, 11(3), 690–6. doi:10.1109/72.846740
- Nie, G., Rowe, W., Zhang, L., Tian, Y., & Shi, Y. (2011). Credit card churn forecasting by logistic regression and decision tree. *Expert Systems with Applications*, 38(12), 15273–15285. doi:10.1016/j.eswa.2011.06.028
- Novakovic, J. (2010). The Impact of Feature Selection on the Accuracy of Naive Bayes Classifier. In *18th Telecommunications forum TELFOR* (Vol. 2, pp. 1113–1116).
- Nugroho, A. S. (2008). Support Vector Machine: Paradigma Baru dalam Softcomputing. *Neural Networks*, 92–99.
- Pan, S., Iplikci, S., Warwick, K., & Aziz, T. Z. (2012). Parkinson's Disease tremor classification – A comparison between Support Vector Machines and neural networks. *Expert Systems with Applications*, 39(12), 10764–10771. doi:10.1016/j.eswa.2012.02.189
- Richeldi, M., & Perrucci, A. (2002). *Churn analysis case study. Telecom Italian Lab*. Torino. Retrieved from [http://www-ai.cs.uni-dortmund.de:8080/PublicPublicationFiles/richeldi\\_perrucci\\_2002b.pdf](http://www-ai.cs.uni-dortmund.de:8080/PublicPublicationFiles/richeldi_perrucci_2002b.pdf)
- Richter, Y., Yom-Tov, E., & Slonim, N. (2010). Predicting customer churn in mobile networks through analysis of social groups. In *Proceedings of the 2010 SIAM International Conference on Data Mining (SDM 2010)* (pp. 732–741). doi:10.1137/1.9781611972801.64
- Rynkiewicz, J. (2012). General bound of overfitting for MLP regression models. *Neurocomputing*, 90, 106–110. doi:10.1016/j.neucom.2011.11.028
- Sharma, A., & Panigrahi, P. (2011). A neural network based approach for predicting customer churn in cellular network services. *International Journal of Computer Applications (0975-8887)*, 27(11), 26–31. doi:10.5120/3344-4605

- Tian, W., Song, J., Li, Z., & de Wilde, P. (2014). Bootstrap techniques for sensitivity analysis and model selection in building thermal performance analysis. *Applied Energy*, 135, 320–328. doi:10.1016/j.apenergy.2014.08.110
- Tsai, C.-F., & Lu, Y.-H. (2009). Customer churn prediction by hybrid neural networks. *Expert Systems with Applications*, 36(10), 12547–12553. doi:10.1016/j.eswa.2009.05.032
- Vapnik, V. (1998). *The Nature of Statistical Learning Theory*. *Technometrics*. John Wiley & Sons, Inc. Retrieved from <http://www.tandfonline.com/doi/pdf/10.1080/00401706.1996.10484565>
- Verbeke, W., Dejaeger, K., Martens, D., Hur, J., & Baensens, B. (2012). New insights into churn prediction in the telecommunication sector: A profit driven data mining approach. *European Journal of Operational Research*, 218(1), 211–229. doi:10.1016/j.ejor.2011.09.031
- Wahono, R. S., Herman, N. S., & Ahmad, S. (2014). Neural Network Parameter Optimization Based on Genetic Algorithm for Software Defect Prediction. *Advanced Science Letters*, 20(10), 1951–1955. doi:10.1166/asl.2014.5641
- Witten, I. H., Frank, E., & Hall, M. A. (2011). *Data Mining Practical Machine Learning Tools and Techniques* (3rd ed.). USA: Morgan Kaufmann Publishers.
- Wu, Q. (2011). Hybrid forecasting model based on support vector machine and particle swarm optimization with adaptive and Cauchy mutation. *Expert Systems with Applications*, 38(8), 9070–9075. doi:10.1016/j.eswa.2010.11.093
- Wu, Xindong & Kumar, V. (2009). *The Top Ten Algorithm in Data Mining*. Boca Raton: Taylor & Francis Group, LLC.
- Xia, G., & Jin, W. (2008). Model of Customer Churn Prediction on Support Vector Machine. *Systems Engineering - Theory & Practice*, 28(1), 71–77. doi:10.1016/S1874-8651(09)60003-X
- Xu, J., Tang, Y. Y., Zou, B., Xu, Z., Li, L., & Lu, Y. (2014). Generalization performance of Gaussian kernels SVM based on Markov sampling. *Neural Networks: The Official Journal of the International Neural Network Society*, 53, 40–51. doi:10.1016/j.neunet.2014.01.013
- Yap, B. W., Rani, K. A., Rahman, H. A., Fong, S., Khairudin, Z., & Abdullah, N. N. (2014). An Application of Oversampling, Undersampling, Bagging and Boosting in Handling Imbalanced Datasets. *Proceedings of the First International Conference on Advanced and Information Engineering*, 285, 429–436. doi:10.1007/978-981-4585-18-7\_2
- Yu, X., Guo, S., Guo, J., & Huang, X. (2011). An extended support vector machine forecasting framework for customer churn in e-commerce. *Expert Systems with Applications*, 38(3), 1425–1430. doi:10.1016/j.eswa.2010.07.049
- Zhou, S.-S., Liu, H.-W., & Ye, F. (2009). Variant of Gaussian kernel and parameter setting method for nonlinear SVM. *Neurocomputing*, 72(13-15), 2931–2937. doi:10.1016/j.neucom.2008.07.016

## BIOGRAFI PENULIS



**Abdul Razak Naufal.** Menyelesaikan pendidikan S1 di Universitas Islam Negeri (UIN) Walisongo, Semarang dan S2 Magister Teknik Informatika di Universitas Dian Nuswantoro Semarang, Indonesia. Saat ini menjadi trainer dan konsultan IT di CV. Media Hasanah. Minat penelitian saat ini adalah data

mining.



**Romi Satria Wahono.** Menempuh pendidikan S1, S2 di departement computer science di Saitama University, Jepang, dan Ph.D di department Software Engineering di Universiti Teknikal Malaysia Melaka. Saat ini menjadi dosen pascasarjana Magister Teknik Informatika di Universitas Dian Nuswantoro, Indonesia. Founder IlmuKomputer.com dan CEO PT Brainmatics, sebuah perusahaan pembuatan software di Indonesia. Minat penelitian saat ini adalah software engineering dan machine learning. Professional member dari ACM and IEEE.



**Abdul Syukur.** Menerima gelar sarjana dibidang Matematika dari Universitas Diponegoro Semarang, gelar master dibidang Manajemen dari Universitas Atma Jaya Yogyakarta, dan gelar doctor dibidang ekonomi dari Universitas Merdeka Malang. Saat ini menjadi dosen dan dekan di Fakultas Ilmu Komputer, Universitas Dian Nuswantoro, Semarang, Indonesia. Minat penelitiannya saat ini meliputi decision support systems dan information management systems.